FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA
NÚCLEO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM DESENVOLVIMENTO REGIONAL E MEIO AMBIENTE

MODELAGEM DE SOFTWARE PARA TOMADA DE DECISÃO NO DESENVOLVIMENTO REGIONAL ATRAVÉS DE INDICADORES

IZAN FABRÍCIO NEVES CALDERARO

Porto Velho (RO)
2016

FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA
NÚCLEO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM DESENVOLVIMENTO REGIONAL E MEIO AMBIENTE

MODELAGEM DE SOFTWARE PARA TOMADA DE DECISÃO NO DESENVOLVIMENTO REGIONAL ATRAVÉS DE INDICADORES

IZAN FABRÍCIO NEVES CALDERARO

Orientador: Prof. Dr. Fabrício Moraes de Almeida

Dissertação de Mestrado apresentada junto ao Programa de Pós-Graduação em Desenvolvimento Regional e Meio Ambiente, Área de Concentração em Políticas Públicas e Desenvolvimento Sustentável para obtenção do Título de Mestre em Desenvolvimento Regional e Meio Ambiente.

Porto Velho (RO)
2016

FICHA CATALOGRÁFICA
BIBLIOTECA PROF. ROBERTO DUARTE PIRES

C146m

Calderaro, Izan Fabrício Neves.

Modelagem de software para tomada de decisão no desenvolvimento regional através de indicadores / Izan Fabrício Neves Calderaro. – Porto Velho, Rondônia, 2016.

78 f.

Orientador: Prof. Dr. Fabrício Moraes de Almeida

Dissertação (Mestrado em Desenvolvimento Regional) - Fundação Universidade Federal de Rondônia – UNIR.

1.Indicadores. 2. Desenvolvimento regional. 3. Modelagem. 4. Teoria Bayesiana. I. Almeida, Fabrício Moraes de. II. Fundação Universidade Federal de Rondônia – UNIR. III. Título.

CDU: 504: 004

Bibliotecária Responsável: Edoneia Sampaio CRB 11/947

IZAN FABRÍCIO NEVES CALDERARO

"MODELAGEM DE SOFTWARE PARA TOMADA DE DECISÃO NO DESENVOLVIMENTO REGIONAL ATRAVÉS DE INDICADORES".

Comissão Examinadora

Dr. Fabrício Moraes de Almeida
Presidente
Fundação Universidade Federal de Rondônia

Dr. Flávio Batista Simão
Membro Externo
Fundação Universidade Federal de Rondônia

Dra. Carolina Yukari Veludo Watanabe
Membro Externo
Fundação Universidade Federal de Rondônia

Dra. Ana Karina Dias Salman
Membro Interno
Fundação Universidade Federal de Rondônia/Embrapa Rondônia

Dra. Carolina Rodrigues da Costa Dória
Suplente
Fundação Universidade Federal de Rondônia

Porto Velho, 8 de Junho de 2016.

Resultado: Aprovado

DEDICATÓRIA

À E. Matteo S. Calderaro o TTO, meu filho, que sempre me fez e faz superar todas as adversidades enfrentadas com apenas um sorriso.

AGRADECIMENTOS

Em primeiro momento agradeço ao meu pai, que sempre me lembrava do porquê estar estudando, apenas pelo simples fato de perguntar o que estava fazendo em casa e não no trabalho.

À minha mãe por ser sempre paciente quando estava sem paciência e me dar palavras de motivação quando a mesma não tinha motivo para tanto.

Aos meus irmãos.

Ao professor (Pós-Doc) Dr. Fabrício Moraes de Almeida por ter me conduzido nesta jornada de desafios e nunca o ter deixado de fazer, mesmo com tantos imprevistos ao longo dos meses que se seguiram.

Às pessoas que tive a oportunidade de conhecer e trocar experiências durante o período do Mestrado.

À Fundação de Amparo ao Desenvolvimento das Ações Científicas e Tecnológicas e a Pesquisa de Rondônia (FAPERO) por incentivar e apoiar este projeto no custeio das atividades.

EPÍGRAFE

O modo como você reúne, administra e usa a informação determina se vencerá ou perderá.

(William Henry Gates III)

RESUMO

A análise e modelagem de dados são necessárias para produzir uma base de dados consistente. Uma base de conhecimento pode ser desenvolvida através do uso de indicadores, que por sua vez são o elo tornando possível a construção de um modelo de dados empregado na tomada de decisão e início da exploração consciente e discriminada dos recursos envolvidos. Dentro do contexto dos sistemas que agem racionalmente, duas abordagens principais podem ser utilizadas: raciocínio lógico e raciocínio probabilístico. O raciocínio lógico pondera sobre o conhecimento prévio a respeito do problema e sobre esta base de conhecimento retira suas conclusões. Esta abordagem, apesar de poderosa, pode não ser útil em situações onde não se conhece previamente todo o escopo do problema, para estes casos, o raciocínio probabilístico surge como uma boa opção. Coloca-se assim um grande desafio às estratégias de desenvolvimento de nações e regiões economicamente subdesenvolvidas, posto que, candidatas como são à inclusão econômica, se deparam com uma condicionante ambiental criado historicamente por outras regiões que já se encontram na dianteira do processo de desenvolvimento. Por outro lado, a nova consciência ambiental que cresce a cada dia, traz também oportunidades que podem e devem ser consideradas nas estratégias de desenvolvimento destas regiões. Assim o objetivo do trabalho foi realizar a análise de indicadores para modelagem de um sistema que forneça subsídios decisivos para a tomada de decisão, um novo tipo de exploração que busque minimizar o impacto ambiental e desenvolver formas mais sustentáveis de produção na utilização dos recursos naturais. Foi usada para isto a metodologia UML, que é uma linguagem de modelagem unificada que permite representar um sistema de forma padronizada, utilizando a Linguagem C para desenvolvimento dos algoritmos e o PostgreSQL como o gerenciador de banco de dados para o armazenamento das variáveis. Com esse objetivo, a dissertação teve como proposta o desenvolvimento da modelagem de uma aplicação, que mensure o grau de importância dos indicadores e os transforme em valores que possam ser utilizados como elementos na tomada de decisão, utilizando para isso um sistema que possa atuar em situações de incerteza. Os resultados estão definidos pelo tratamento dos indicadores pela teoria Bayesiana e sua análise.

Palavras-chaves: Indicadores, Desenvolvimento Regional, Modelagem, Teoria Bayesiana.

ABSTRACT

The modeling and analysis of data is necessary to produce a consistent database, a knowledge base can be developed through the use of indicators, which in turn is the link making it possible to build a data model used in the decision-making and the beginning of conscious and discriminated exploitation of resources involved. Within the context of systems that act rationally, two main approaches can be used: logical reasoning and probabilistic reasoning. The logical reasoning ponders prior knowledge about the problem and on this basis of knowledge draws its conclusions. This approach, although powerful, can not be useful in situations where not previously know the whole scope of the problem, for these cases, probabilistic reasoning comes as a good option. This raises a major challenge to development strategies of nations and economically underdeveloped regions, since, candidates as are the economic inclusion, are faced with an environmental condition created historically by other regions which are already at the forefront of the development process. On the other hand, the new environmental awareness that is growing every day, it also brings opportunities that can and should be considered in the development strategies of these regions. Thus the aim of the study was the analysis of indicators for modeling a system that provides decisive benefits for decision-making, a new type of exploitation that seeks to minimize the environmental impact and develop more sustainable forms of production in the use of natural resources. Was used for this UML methodology, which is a unified modeling language that allows represent a standardized way system, using the C language for the development of algorithms and PostgreSQL as the database manager for storing variables. For this purpose, the dissertation is proposed the development of modeling an application that measures the degree of importance of indicators and transform them into values that can be used as elements in decision-making, making use of a system that can work in situations uncertainty. The results are set for processing the indicators by Bayesian theory and analysis.

Keywords: Indicators, Regional Development, Modeling, Bayesian Theory.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE EQUAÇÕES

LISTAS DE SIGLAS E ABREVIATURAS

ACID – Atomicity, Consistency, Isolation, Durability.

API – Application Programming Interface.

BUGS – Bayesian Inference Using Gibbs Sampling.

BAYESX – Bayesian Inference in Structured Additive Regression Models.

CRAN – The Comprehensive R Archive Network.

DAG – Directed Acyclic Graph.

DBA – Data Base Administrator.

DBI – Data Base Interface.

DDL – Data Definition Language.

DML – Data Manipulation Language.

GAM – Generalized Additive Models.

GAMM – Generalized Additive Mixed Models.

GGAMM – Generalized Geoadditive Mixed Models.

GPLv3 – General Public License version 3.

GUI - Graphic User Interface.

IDAM – Integrative Dam Assessment Modeling.

IDE – Integrated Development Environment.

IDH – Indice de Desenvolvimento Humano.

INLA – Integrated Nested Laplace Approximations.

JUDE – Java and UML Developers Environment.

JAGS – Just Another Gibbs Sampler.

JOINS – Junções de Tuplas em Banco De Dados.

MCMC – Markov Chain Monte Carlo.

MRBAYES – Bayesian estimation of phylogeny.

PNDR – Plano Nacional de Desenvolvimento Regional.

PRODES – Monitoramento da Floresta Amazônica Brasileira por Satélite.

RB – Rede Bayesiana.

SQL – Structure Query Language.

SGDB – Sistemas Gerenciadores de Banco de Dados

UML – Unified Modeling Language.

UDF – User-Defined Function

WWW – World Wide Web.

wxWidgets – Cross-Platform GUI Library.

SUMÁRIO

INTRODUÇÃO

A presente dissertação tem como finalidade tratar da modelagem de indicadores com o propósito de construir uma rede bayesiana para utilização como ferramenta de auxílio na tomada de decisão. A modelagem de sistema apresentada aqui procura ser simples e ter sua base voltada para a análise dos indicadores a serem utilizados, com a possibilidade de confecção de um modelo geral. Além disso, é uma modelagem adaptativa.

Neste trabalho será abordado um enfoque fundamental para a construção de um sistema capaz de realizar a inferência e o aprendizado em redes bayesianas. “Redes Bayesianas são grafos acíclicos dirigidos que representam dependência entre variáveis em um modelo probabilístico” (MARQUES; DUTRA, 2013). Essa abordagem representa uma estratégia para resolver os problemas que tratam de incertezas. A tarefa mais comum realizada para a utilização da rede bayesiana é determinar várias probabilidades que são relevantes, condicionadas a certos eventos observados.

Estas probabilidades não são armazenadas diretamente no modelo, e consequentemente precisam ser computadas por algoritmos de inferência. Outra questão importante nessa abordagem é que a construção manual de uma rede bayesiana pode ser um processo bastante trabalhoso e caro. Por isso este trabalho irá tentar ser o mais simples possível, na modelagem do sistema, que no futuro poderá ser construído com base nesta análise.

Além disso, desenvolver uma aplicação que mensure o grau de importância dos indicadores e os transforme em valores que possam ser utilizados como elementos na tomada de decisão, utilizando para isso “um sistema que possa atuar em situações de incerteza, capaz de atribuir níveis de confiabilidade para todas as sentenças” (RUSSEL; NORVIG, 2004) em sua base de conhecimento, e ainda, estabelecer relações entre as sentenças.

Como objetivo geral será mostrado o desenvolvimento da modelagem de uma aplicação que mensure o grau de importância dos indicadores e os transforme em valores que possam ser utilizados como elementos na tomada de decisão para análise do desmatamento, utilizando para isso os diagramas da UML para representar as diversas faces do sistema que possa atuar em situações de incerteza, capaz de atribuir níveis de confiabilidade para todas as sentenças em sua base de conhecimento, e ainda, estabelecer relações entre as sentenças.

Já como objetivos específicos estão definidos:

- Investigar os indicadores que serão utilizados para compor a base de dados;
- Desenvolver a modelagem de software através dos diagramas UML representando a atuação de Redes Bayesianas com os indicadores;

Procuro definir uma modelagem de software para a tomada de decisões, baseadas em crenças, definindo o grau de incerteza, ingenuidade heurística e um classificador probabilístico simples baseado na aplicação de teorema de Bayes, com fortes hipóteses de independência entre seus atributos e verossimilhança, isto é, um nível de conhecimento elevado à possibilidade e inferior à probabilidade.

Hoje, a sociedade não aceita mais um modelo de crescimento de exploração indiscriminada de recursos naturais que comprometa o planeta para as gerações futuras.

Assim, de acordo com o Brasil (2012), não há mais lugar para elaboração de políticas de desenvolvimento setorial e espacial, urbano e regional, sem que se considerem, simultaneamente, a sustentabilidade social e ambiental. Tal restrição passa a exigir padrões diferentes de consumo com muito talento e racionalidade nos usos dos recursos naturais, especialmente da água e de fontes geradoras de energia, da mesma forma, como também no destino final dos resíduos. Longe deste objetivo, apesar de ainda ser um plano, é o pontapé inicial para a conscientização da sociedade em procurar maneiras mais adequadas de políticas públicas para um melhor aproveitamento de nossos recursos naturais, prejudicando ao mínimo possível, em último caso, o ambiente.

Segundo o Brasil (2012), apresenta-se um grande desafio às estratégias de desenvolvimento de nações e regiões economicamente subdesenvolvidas, posto que, candidatas como são à inclusão econômica, se deparam com um condicionante ambiental criado historicamente por outras regiões que já se encontram na dianteira do processo de desenvolvimento. Por outro lado, a nova consciência ambiental que cresce a cada dia, traz também oportunidades que podem e devem ser consideradas nas estratégias de desenvolvimento destas regiões.

Possivelmente, a consciência ecológica não aconteceu com os empreendimentos que envolvem a coleta e o aproveitamento dos meios naturais, não ocorre com as consequências descritas previamente e estabelecidas nos projetos iniciais, pelo qual quais enormes discrepâncias são observadas entre o pré-projeto e pós-projeto. Por exemplo, a construção das recentes Usinas do Madeira: UHE Santo Antônio no município de Porto Velho e UHE Jirau no distrito de Jaci-Paraná.

Em um artigo recente da Energy Policy, escrito por pesquisadores da Universidade de Oxford (ANSAR et al., 2014) foram analisadas 245 hidrelétricas construídas em 65 países entre 1934 e 2007 e afirmou que, na maioria dos casos, as grandes usinas hidrelétricas não são viáveis economicamente. E concluído que elas ultrapassam os orçamentos, afogam a economia dos países e não entregam os benefícios prometidos.

A inovação envolve conceitos que levam inevitavelmente à complexidade dos processos de Desenvolvimento, por ser subjacente a ele e ter uma natureza igualmente característica de diversidade e integração. Corresponde à criação de meio sobre o que já existe ou não. Quanto maior sua condição de inovação se difere da invenção simples e predispõe à imitação. Está no ato de criação, no processo de desenvolvimento e no produto resultante. “Deve servir à rentabilidade econômica das empresas e ainda mais à melhoria da qualidade de vida das pessoas, com o viés da sustentabilidade” (FRANZIN; ALMEIDA; SOUZA, 2014).

Dentro do contexto dos sistemas que agem racionalmente, duas abordagens principais podem ser utilizadas: raciocínio lógico e raciocínio probabilístico. O raciocínio lógico pondera sobre o conhecimento prévio a respeito do problema e sobre esta base de conhecimento retira suas conclusões. Esta abordagem, apesar de poderosa, pode não ser útil em situações onde não se conhece previamente todo o escopo do problema, para estes casos, o raciocínio probabilístico surge como uma boa opção. “A principal vantagem de raciocínio probabilístico sobre o raciocínio lógico é o fato de que agentes podem tomar decisões racionais mesmo quando não existe informação suficiente para se provar que uma ação funcionará” (CHARNIAK, 1991).

Segundo Pereira (2008), as Redes Bayesianas vêm sendo utilizadas para a solução de vários tipos de problemas em diversas áreas, como por exemplo: diagnóstico médico, aprendizado de mapas, interpretação de linguagem e visão.

Para Habrant (1999), as Redes Bayesianas apresentam a possibilidade de realizar a modelagem da base de conhecimento especialista de forma automatizada através da análise de base de dados. Quando uma Rede Bayesiana está ligada á uma base de dados ela pode ser construída de forma a se tornar adaptativa e se atualizar conforme as possibilidades estimadas a partir dos dados armazenados.

As Redes Bayesianas oferecem uma abordagem para raciocínio probabilístico que engloba a teoria de grafos para o estabelecimento das relações entre as sentenças e ainda a teoria de probabilidades, para a atribuição de níveis de confiabilidade. Neste sentido questiona-se:

Quais decisões serão adequadas a serem tomadas quando são utilizados indicadores do Desenvolvimento Regional e Sustentabilidade a fim de determinar o desflorestamento na

Amazônia? A hipótese adotada é de que a conversão de áreas de florestas para uso agropecuário, ou seja, pastagens e lavouras é um fator determinante do desmatamento entre outras variáveis que podem ser acrescentadas.

Partindo deste pressuposto seria mais adequado abordar um processo mais conciente tomando como base empreendimentos passados, políticas públicas, inovação e redes Bayesianas.

No capítulo 1 é mostrado o referencial teórico, do conhecimento, empregado no desenvolimento desta dissertação, apresentando a importância dos indicadores para o desenvolvimento regional e sustentável, os objetivos e conceitos dos indicadores, a modelagem integrativa para avaliação de barragens (IDAM), desenvolvimento regional e sustentável, princípio da incerteza, inferência Bayesiana, linguagem de modelagem unificada (UML), PostgreSQL, estrutura léxica SQL, A Linguagem C e suas funções definidas pelo usuário (UDF), além de como é definida suas estruturas.

Já no capítulo 2 são apresentados os materiais e métodos que poderão a vir ser usados nesta modelagem. Será mostrado um método quantitativo e o que vem a ser um método qualitativo, o emprego da metodologia escolhida é definido neste assim como a apresentação dos materiais utilizados, como o aplicativo para a modelagem UML, para a construção dos algoritmos na Linguagem C, banco de dados, gerenciador de banco de daos e sistema operacional que por ventura poderá ser utilizado como plataforma para a aplicação da solução.

O capítulo 3 apresenta a definição dos indicadores, o tratamento quantitativo destes indicadores assim como a definição das variáveis que podem ou não influenciar os mesmos, a computação Bayesiana que envolve o processo heurístico também foi apresentada assim como algumas aplicações existentes no mercado, programas estes de código aberto e proprietários de código fechado.

No capítulo 4, aborda-se a modelagem de software para a tomada de decisão é descrita na íntegra com a heurística e computação Bayesiana analisada através da UML, apresentando uma modelagem dos indicadores assim como a estrutura que possa ser construída no Sistema Gerenciador de Banco de Dados (SGDB) para seu armazenamento. São apresentadas as funções que devem ser desenvolvidas para atender o modelo computacional Bayesiano.

No capítulo 5 algumas questões são levantadas “de o por que” da necessidade da análise de sistema para o desenvolvimento de um modelo para qualquer aplicação, por mais simples que o objetivo desta possa ser.

Por último, as considerações finais e a proposta deixada para trabalhos futuros seguindo a mesma linha de pesquisa apresentada nesta dissertação.

CAPÍTULO 1: REFERENCIAL TEÓRICO

Segundo Bellen (2006), os modelos de indicadores de sustentabilidade são usados para traçar um modelo da realidade, avaliar condições e tendências, comparar situações e lugares, avaliar metas e objetivos, antecipar futuras condições e tendências.

Os indicadores são instrumentos de gestão essenciais nas atividades de monitoramento e avaliação, assim como seus projetos, programas e políticas, pois permitem acompanhar o alcance das metas, identificar avanços, melhorias de qualidade, correção de problemas, necessidades de mudança conforme Mpog (2009).

Pode-se dizer que os indicadores possuem, minimamente, duas funções básicas:

- A primeira é descrever por meio da geração de informações o estado real dos acontecimentos e o seu comportamento;
- A segunda é de caráter valorativo que consiste em analisar as informações presentes com base nas anteriores de forma a realizar proposições valorativas.

1.1 A IMPORTÂNCIA DA DEFINIÇÃO DE INDICADORES PARA O DESENVOLVIMENTO REGIONAL SUSTENTÁVEL

Para Prescott-Allen (1999), a efetivação do desenvolvimento sustentável caracteriza-se por meio de uma economia robusta, sistemas naturais ricos e flexíveis e comunidades prósperas. Mas para o alcance deste desenvolvimento é necessário planejamento e principalmente monitoramento. E nessa lógica, os indicadores de sustentabilidade minimizam as chances de se obter resultados não pretendidos.

Para Moldan e Bilharz (1997), decisões são tomadas dentro de todas as esferas da sociedade e são influenciadas por valores, tradições e por uma série de *inputs* de várias direções. A efetividade e a racionalidade do processo podem ser incrementadas pelo uso apropriado da informação, e os indicadores podem ajudar no processo decisório.

Como Bakkes (1994) afirma diversos passos podem ser identificados para o processo de tomada de decisão no contexto da sustentabilidade e de seus indicadores; identificação do problema, desenvolvimento de política e controle.

Para Bellen (2006) modelos de indicadores de sustentabilidade são usados para traçar um modelo da realidade, avaliar condições e tendências, comparar situações e lugares, avaliar metas e objetivos, antecipar futuras condições e tendências.

Portanto, uma estrutura analítica e bem elaborada de indicadores permite integrar, de maneira concisa, informações de cunho social, ecológico, econômico e geográficos, com graus de importância distintos. Dessa forma é possível analisar se as diretrizes estabelecidas na política pública estão alcançando o desenvolvimento e tendo o resultado esperado e quais são os fatores essencialmente responsáveis por este sucesso, permitindo agir sobre esses fatores. Potencializar resultados positivos contribui para estabelecer resposta rápida na busca por melhores condições de vida da população.

1.2 CONCEITO DOS INDICADORES

O indicador: "é a medida do comportamento do sistema em termo de atributos expressivos e perceptíveis [...]" de modo a estabelecer "[...] um parâmetro ou valor derivado de parâmetros que apontam e fornecem informações sobre o estado de um fenômeno [...]" (BELLEN, 2006).

Por meio de indicadores bem elaborados, é possível traçar metas e objetivos e mensurar o atingimento dos resultados de maneira clara e transparente, evitando armadilhas e desvirtuamento de planejamento. Segundo Hardi e Barg (1997), tais indicadores servem para identificar variações, comportamentos, processos e tendências; estabelecer comparações entre países e entre regiões; indicar necessidades e prioridades para a formulação, monitoramento e avaliação de políticas; e, por sua capacidade de síntese, são capazes de facilitar o entendimento ao crescente público envolvido com o tema.

Os indicadores são utilizados para:

- Mensurar os resultados e gerir o desempenho;
- Embasar a análise crítica dos resultados obtidos e do processo de tomada decisão;
- Contribuir para a melhoria contínua dos processos organizacionais;
- Facilitar o planejamento e o controle do desempenho;
- Viabilizar a análise comparativa do desempenho de organizações.

Pesquisadores e institutos de planejamento, pelo mundo, têm empreendido esforços para o desenvolvimento de modelos de indicadores voltados ao desenvolvimento sustentável, Bellen (2006) seleciona em seu estudo 18 deles, sendo os mais conhecidos: Pegada Ecológica (*Ecological Footprint*), Barômetro da Sustentabilidade (*Barometer of Sustainability*) e o Painel de Sustentabilidade (*Dashboard of Sustainability*).

Para Hardi (2000), o Painel de Sustentabilidade possui como destaque em relação aos demais modelos, a investigação não somente de cada dimensão envolvida, mas também como estas dimensões interagem para determinar a sustentabilidade do sistema. O quadro 1 sintetiza o conceito de indicadores descrito por diversos autores.

Quadro 1 – Conceito de indicadores. Fonte: Bellen (2006). Adaptado por Kramma (2009).

CONCEITO	AUTOR
O termo indicador é originário do latim *indicare*, que significa descobrir, apontar, anunciar, estimar.	Hammond et al. (1995)
Para conceituar indicadores é necessário alcançar maior clareza e consenso nessa área, tanto em relação à definição de indicadores quanto a outros conceitos associados como: índice, meta e padrão.	akkes et al. (1994)
Um indicador é uma medida que resume informações relevantes de um fenômeno particular ou um substituto dessa medida.	McQueen e Noak (1988)
Um indicador deve ser entendido como um parâmetro, ou valor derivado de parâmetros que apontam e fornecem informações sobre o estado de um fenômeno, com uma extensão significativa.	OCDE (1993)
É uma medida do comportamento do sistema em termos de atributos expressivos e perceptíveis.	Holling (1978)
Indicador é uma ferramenta que permite a obtenção de informações de uma dada realidade.	Mitchel (1996)
Um indicador pode ser um dado ou um agregado de informações, sendo que um bom indicador deve conter os seguintes atributos: simples de entender; quantificação estatística e lógica coerente; e comunicar eficientemente o estado do fenômeno observado.	Mueller et. al. (1997)
Os indicadores podem comunicar ou informar sobre o progresso em direção a uma determinada meta, mas também podem ser entendidos como um recurso que deixa mais perceptível uma tendência ou fenômeno que não seja imediatamente detectável.	Hammond et. al. (1995)
Os indicadores, em nível mais concreto, devem ser entendidos como variáveis.	Gallopin (1996)
Algumas definições colocam um indicador como uma variável que está relacionada hipoteticamente com outra variável estudada, que não pode ser diretamente observada.	Chevalier et al. (1992)
Indicadores são sinais referentes a eventos e sistemas complexos. São pedaços de informação que apontam para características dos sistemas, realçando o que está acontecendo.	Hardi e Barg (1997)

1.3 MODELAGEM INTEGRATIVA PARA AVALIAÇÃO DE BARRAGENS

O Idam (2012) fornece uma representação visual dos vários custos e benefícios associados a duas ou mais represas, esta ferramenta permite aos tomadores de decisão avaliar alternativas e articular as prioridades associadas a um projeto de represa, tornando o processo de decisão sobre ela mais informada e mais transparente.

Cada um dos vinte e um dos diferentes impactos utilizados na construção de uma represa são avaliados objetivamente (por exemplo, enchente e proteção) e subjetivamente (ou seja, a avaliação da referida proteção contra as inundações).

> A ferramenta de Modelagem Integrativa para Avaliação de Barragens do inglês *Integrative Dam Assessment Modeling* (IDAM) é projetada para integrar perspectivas biofísicas, socioeconômicas e geopolíticas em uma análise de custo-benefício único na construção de barragens (IDAM, 2012).

Neste modelo para todos os indicadores são definidos impactos positivos e negativos, apresentando o ponto de partida, ou seja, por meio da análise de indicadores pode-se definir um modelo para a tomada de decisão, que neste caso está definido para a construção consciente de represas, ressaltando-se que a "consciência" não é o forte dos empreendimentos de geração de energia usando recursos naturais hídricos. O IDAM surgiu para minimizar os impactos negativos sobre o meio ambiente.

1.4 DESENVOLVIMENTO REGIONAL E SUSTENTÁVEL

A concepção de "Desenvolvimento" passa por uma série de campos teóricos e práticos que não se confundem entre si, como o da História, da Sociologia e da Economia, mas é importante destacar de início a especialização do termo.

Desenvolvimento, em princípio, é genericamente o que está em movimento, em processo. Restritamente, consiste em uma noção de avanço e progresso, no sentido de atingir continuamente pontos de superação de resultados, em uma cadeia de inter-relações e de forças propulsoras.

Assim, segundo Franzin; Almeida; Souza (2014), nem tudo que está em Desenvolvimento enquanto processo é Desenvolvimento enquanto avanço das condições de subsistência e melhoria de vida. Desenvolvimento Regional é, portanto uma expressão que está para além do que se produz ou se faz nas rotinas cotidianas como ações em movimento nas rodas das vivências coletivas.

É importante antecipar também ressaltar que “Desenvolvimento” não se limita a lucratividade. Para Franzin; Almeida; Souza (2014), expectativa de reduzir todo um conjunto de ações e resultados aos objetivos da atividade econômica constrói ou alimenta a ideologia do capitalismo alienado, em que fatores de sustentação ou de risco são tomados como elementos insignificantes, tornando insignificante a própria condição humana. Expressamos claramente que o capitalismo não é alienado e sim alienador.

Segundo North (1977), que aplicou estudos sobre a industrialização nos Estados Unidos e fez um comparativo com outros países ou regiões de modelo capitalista e que não sofreram “[...] restrições impostas por pressão populacional”, o desenvolvimento regional se faz ideologicamente dentro de algumas fases, nesta ordem: economia de subsistência, industrialização modesta, sucessão de culturas agrícolas, industrialização forçada e especialização em atividades terciárias com produtos de exportação. O contexto justifica, nos termos de North (1977), que “[...] uma teoria do crescimento econômico regional deveria, claramente, concentrar-se nos fatores críticos que promovem ou impedem o desenvolvimento”. Não se limita à industrialização, North (1977) afirma "se desejarmos um modelo normativo de como as regiões deveriam crescer, com o objetivo de analisar as causas da estagnação ou decadência, então, essa sequência de estágios é de pouca utilidade e de fato enganadora, pela ênfase que coloca na necessidade da industrialização (e nas dificuldades de promovê-la)".

Os problemas ambientais são desafios para as ciências econômicas e seu instrumental analítico precisa ser capaz de fornecer respostas consistentes para harmonizar as relações entre meio ambiente e economia.

Para Oliveira (2012), sociedade e as atividades econômicas são, inevitavelmente, dependentes dos bens e serviços fornecidos pelo meio ambientes de modo que é fundamental que a teoria econômica considere as interconexões entre sistema econômico e ambiente externo.

> A questão ambiental e o problema do desenvolvimento sustentável não se restringem apenas à esfera ambiental, mas incluem no debate questões relacionadas às estruturas social e econômica, as quais envolvem elevado grau de complexidade, incerteza e desconhecimento, além de um imperativo ético relacionado ao tratamento das gerações futuras pelas gerações presentes (OLIVEIRA, 2012).

Observa-se que a intenção é em primeiro lugar, segundo Amazonas (2001), contestar a ideia de que os limites biofísicos ambientais se constituem em limites ao crescimento econômico, defendendo que as inovações tecnológicas induzidas pela escassez dos recursos ambientais tendem a superar as restrições colocadas por estes. O outro propósito é propor que os danos ambientais sejam entendidos em termos dos custos sociais, que devem ser internaliza-

dos. Sendo assim, enquanto a economia dos recursos naturais esvazia as preocupações com a justiça e equidade sociais quando pressupõe que o progresso técnico e a substituição entre recursos sejam capazes de superar tais restrições, a economia da poluição propõe que os custos ambientais sejam internalizados por meio da valoração dos recursos ambientais. Desse modo, as duas abordagens neoclássicas, apesar de distintas, são complementares em seu propósito.

> Os primeiros trabalhos foram publicados no início dos anos 1960. No ano de 1963, Barnett e Morse publicaram *Scarcity and Growth: The Economics of Natural Resource Availability*, um trabalho seminal. No ano anterior, havia sido publicada uma análise estatística da evolução histórica dos preços dos recursos por Christy e Potter (OLIVEIRA, 2012).

A economia neoclássica, por ser hegemônica, vem se destacando no tratamento dada à problemática ambiental. Suas formulações teóricas têm por princípio o individualismo metodológico, o utilitarismo e o equilíbrio, desenvolvendo uma compreensão do sistema econômico como sendo constituído, basicamente de indivíduos que se comportam com base em uma racionalidade maximizadora do bem-estar, o que conduz a um resultado ótimo numa situação de equilíbrio. A economia ambiental neoclássica, de acordo com Pearce (2002), surge com a criação da instituição americana *Resources for the Future* (RFF) que procurava utilizar as teorias e instrumentos da economia para estudar as questões ambientais. Seu foco inicial era a escassez de recursos naturais.

> Nessa década, ocorre a primeira revolução ambiental com a publicação de *Silent Spring*, por Rachel Carson, em 1969, na qual procurava alertar sobre os efeitos externos dos agrotóxicos no meio ambiente. Os economistas passaram a utilizar o conceito de externalidades para interpretar as crescentes preocupações ambientais (PEARCE, 2002).

1.5 GRAFOS

A teoria dos grafos é um ramo da matemática que estuda as relações entre os objetos de determinado conjunto. De uma forma simplista um grafo constitui um conjunto, este por sua vez é formado por pontos ligados por linhas (JENSEN, 2001).

Formalmente, um par de conjuntos (V, E) em que:

- V constitui um conjunto não vazio cujos elementos são chamados vértices ou nós;

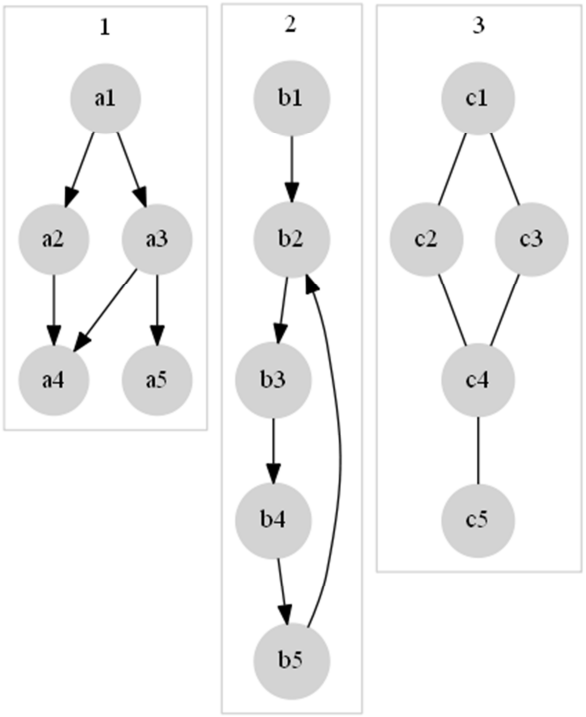
- *E* consiste em um conjunto de pares dos vértices de V, cujos elementos são denominados arestas.

Os conjuntos *V* e *E*, geralmente, são conjuntos finitos:

$$\mathrm{V} = \{v_1, \ldots, v_n\} \text{ e } \mathrm{E} = \{(v_i, v_j), v_i \in \mathrm{V}, v_j \in \mathrm{V}, 0 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{N}\} \quad (1)$$

Para a correta compreensão de uma Rede Bayesiana, é importante entender o que são grafos direcionados e não direcionados. Se as arestas são constituídas de pares ordenados de vértices, diz-se que o grafo é direcionado.

Figura 1 – Grafo direcionado acíclico (1), grafo direcionado cíclico (2) e grafo não direcionado (3).



Em um grafo direcionado, se existe uma aresta de v_i para v_j, diz-se que v_i é pai de v_j. Se existe um caminho direcionado de v_i para v_j diz-se que v_i é ancestral de v_j. De acordo com Russel e Norvig (2004) se um grafo direcionado não possuir ciclos direcionados, isto é, se, para qualquer vértice $v_i \in V$, não existe um caminho direcionado que começa e termina em v_i, então diz-se que é um grafo acíclico direcionado - DAG (*Directed Acyclic Graph*).

Uma Rede Bayesiana fica bem representada por um grafo direcionado acíclico devido ao fato de que ao surgir uma nova evidência, todo o modelo tem que ser executado do início até o seu fim.

1.6 REDES BAYESIANAS

De acordo com Charniak (1991), a melhor maneira de entender as Redes Bayesianas consiste em imaginar-se tentando modelar uma situação em que a casualidade desempenha papel importante, mas em que a compreensão do que está realmente acontecendo é incompleta. Assim, precisa-se descrever a situação de forma probabilística.

A Rede Bayesiana pode ser definida como segundo Jensen (2001):

- Um conjunto de variáveis e um conjunto de arestas direcionadas entre as variáveis;
- Cada variável tem estados finitos e mutuamente exclusivos;
- As variáveis e as arestas direcionadas representam um grafo acíclico direcionado (DAG);
- Cada variável A, com pais $B_1, B_2, \ldots B_n$, possui uma tabela de probabilidades condicionais $\mathrm{P}(\mathrm{A}|B_i..B_n)$, associada.

Para Neapolitan et al. (2004), as Redes Bayesianas são estruturas gráficas para representar as relações probabilísticas entre um grande número de variáveis e para fazer inferência estatística com essas variáveis.

> As Redes Bayesianas permitem eficiente e efetiva representação da distribuição da probabilidade conjunta sobre um grupo de variáveis aleatórias. O objetivo dos modelos gráficos probabilísticos é criar uma estrutura matemática que une grafos e probabilidades e que permita modelar situações complexas envolvendo aleatoriedade ou incerteza (NEAPOLITAN ET AL., 2004).

"A regra da cadeia vale tanto para distribuições condicionais discretas quanto para contínuas se a condição de Markov é satisfeita. RB com variáveis discretas satisfazem a condição de Markov" (NEAPOLITAN et al., 2004), que é dada por: cada variável da Rede Bayesiana é condicionalmente independente do conjunto de todos os não-descendentes desta, dado o conjunto de todos os seus pais.

Em uma Rede Bayesiana, a distribuição conjunta de probabilidades de um conjunto de variáveis discretas, $\{X_1, X_2, \ldots, X_n\}$, é igual ao produtório das distribuições condicionais de todos os nós, dados os valores dos seus pais, ou seja, é dada pela regra da cadeia:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=0}^{n} P(X_i | Pa_i) \tag{2}$$

Os parâmetros de uma Rede Bayesiana são definidos como:

$$\Theta_i = P(X_i | Pa_i), i = 1, \ldots, n \tag{3}$$

Em que, Θ_i é uma tabela de probabilidades condicionais de X_i dado seus pais Pa_i.

Com isso, o conjunto de parâmetros de uma Rede Bayesiana é dado por $\Theta_i = \{\Theta_1, \Theta_2, \ldots, \Theta_n\}$ ou seja, todas as tabelas de probabilidades condicionais da RB com variáveis discretas $\{X_1, X_2, \ldots, X_n\}$

Um importante aspecto de uma Rede Bayesiana refere-se à sua estrutura (topologia do grafo), a qual permite a representação de complexas relações entre variáveis de forma gráfica e intuitiva. A estrutura gráfica de uma Rede Bayesiana facilita o entendimento das relações entre variáveis do seu domínio, além de permitir o uso combinado de informações obtidas do conhecimento de especialistas e de dados históricos para obter a distribuição conjunta de probabilidades da rede.

1.7 INFERÊNCIA BAYESIANA

Existem algumas dificuldades na utilização do modelo Bayesiano conforme Berg e Insua (1996) são elas: a escolha da distribuição a priori e a computação do modelo escolhido. Porém, a escolha da distribuição a priori é considerada o maior problema. Este pode, bem, ser o caso em que o conhecimento subjetivo sobre os parâmetros desconhecidos é avaliado e pode ser incorporado à subjetividade própria das densidades a priori para estes parâmetros. Isto é claramente desejável se puder ser realizado. Algumas ferramentas computacionais recentes tem permitido a aplicação de métodos bayesianos para modelos de alta complexidade e não padronizados. Na verdade, para modelos mais complicados, a análise bayesiana tenha talvez, se tornado o mais simples e frequentemente o único, método de análise.

Segundo Berg e Insua (1996) a base da aplicação da teoria bayesiana é conceitual e prática: fornece uma estrutura coerente e facilita a análise de problemas de decisão sobre incertezas. As críticas relativas aos métodos bayesianos estão centradas em três aspectos: computacionais, imprecisão e descritiva.

Inferência estatística refere-se à obtenção de conclusões sobre quantidades não observadas θ a partir de dados observados y. A inferência Bayesiana aborda o problema definindo a probabilidade de uma forma subjetiva, como uma medida da plausibilidade de uma proposição, condicional no conhecimento do observador. A incerteza em relação a θ pode assumir diferentes graus, os quais se representam através de modelos probabilísticos para θ. Portanto, tanto as quantidades observáveis, quanto os parâmetros do modelo estatístico são considerados quantidades aleatórias.

> Esta última característica constitui uma diferença fundamental da abordagem Bayesiana com a clássica, que considera o parâmetro como uma quantidade fixa e desconhecida, à qual nos aproximamos no processo de inferência (BERNARDO E SMITH 1994).

"Desde o ponto de vista prático, a modelagem Bayesiana inicia com a especificação de um modelo probabilístico completo, através da distribuição das quantidades observáveis e não observáveis do problema" (GELMAN *ET AL*, 1995). A informação disponível sobre θ, resumida na densidade de probabilidade $p(\theta)$, é aumentada com a observação de uma quantidade aleatória y que se relaciona com θ. O teorema de Bayes fornece a regra de atualização desta informação.

De acordo com Carlin e Louis (2000), o modelo Bayesiano mais básico tem dois estágios, com uma especificação de verosimilhança $f(y|\theta)$ e uma especificação anterior $\pi(\theta)$ onde ambas podem ser vetores. Nesta análise Bayesiana mais simples π é assumida como conhecida, de modo que o cálculo da probabilidade e a distribuição posterior são dadas a seguir:

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)} \tag{4}$$

onde:

$$m(y) = \int f(y|\theta)\pi(\theta)d(\theta) \tag{5}$$

A densidade marginal dos dados $f(y|\theta)$ é um caso especial do Teorema de Bayes, esta avaliação de integrais costumava ser difícil ou impossível forçando o método bayesiano em aproximação pouco atraente. "No entanto com o desenvolvimento recente de métodos computacionais de Monte Carlo permitiu uma estimativa precisa do valor de tais integrais, possibilitando a análise avançada de dados Bayesiana" (CARLIN; LOUIS, 2000).

Considere uma quantidade de interesse desconhecida θ (tipicamente não observável). A informação de que dispomos sobre θ, resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ. A distribuição de amostras $p(x|\theta)$ define esta relação.

"A idéia de que após observar $X = x$ a quantidade de informação sobre θ aumentar é bastante intuitiva e o teorema de bayes é a regra de atualização utilizada para quantificar este aumento de informação" (CARLIN; LOUIS, 2000).

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta} \tag{6}$$

Note que $1/p(x)$, que não depende de θ, funciona como uma constante normalizadora de $p(\theta|x)$.

Para um valor de x, a função $l(\theta; x) = p(x|\theta)$ fornece a *plausibilidade* ou *verosimilhança* de cada um dos possíveis valores de θ enquanto $p(\theta)$ é chamada distribuição a *priori* de . Estas duas fontes de informação, priori e verossimilhança, são combinadas levando à distribuição a *posteriori* de θ, $p(\theta|x)$. Assim de acordo com Carlin e Louis (2000), a forma usual de bayes é:

$$p(\theta|x) \propto l(\theta; x)p(\theta) \tag{7}$$

Em palavras temos que;

$$distribuição\ a\ posteriori\ \propto\ verossimil\ \ ança\ x\ distribuição\ a\ priori$$

Note que, ao omitir o termo $p(x)$, a igualdade na equação (3) foi substituída por uma proporcionalidade. Esta forma simplificada do teorema de Bayes será útil em problemas que envolvem estimação de parâmetros já que o denominador é apenas uma constante normaliza-

dora. Em outras situações, como seleção de modelos, este termo tem um papel crucial segundo (CARLIN; LOUIS, 2000).

É intuitivo também que a probabilidade a *posteriori* de um particular conjunto de valores de θ será pequena se $p(\theta)\ ou\ l(\theta; x)$ for pequena para este conjunto. Em particular, se atribuirmos probabilidade a *priori* igual a zero para um conjunto de valores de θ então a probabilidade a *posteriori* será zero qualquer que seja a amostra observada.

A constante normalizadora da *posteriori* pode ser facilmente recuperada, pois $p(\theta|x) = kp(x|\theta)p(\theta)$ onde;

$$k^{-1} = \int f(y|\theta)p(\theta)d(\theta) = E_{\theta|x}[p(X|\theta)] = p(x) \tag{8}$$

chamada distribuição preditiva. Esta é a distribuição esperada para a observação $x\ dado\ \theta$. Assim,

- Antes de observar X pode-se checar a adequação da *priori* fazendo predição via $p(x)$;
- Se X observado recebia pouca probabilidade preditiva então o modelo deve ser questionado.

Se, após observar $X = x$, estamos interessados na previsão de uma quantidade Y, também relacionada com θ, e descrita probabilisticamente por $p(y|\theta)$, então:

$$p(y|x) = \int p(y, \theta|x)d\theta \int p(y|\theta, x)p(\theta|x)d\theta = \int p(y|\theta)p(\theta|x)d\theta \tag{9}$$

onde a última igualdade se deve a independência entre $X\ e\ Y$ condicionado em θ. Esta hipótese de independência condicional está presente em muitos problemas estatísticos. Note que as previsões são sempre verificáveis uma vez que Y é uma quantidade observável. Finalmente, segue da equação 6 que;

$$p(y|x) = E_{\theta|x}[p(X|\theta)] \tag{10}$$

Fica claro também que os conceitos de *priori* e *posteriori* são relativos à aquela observação que está sendo considerada no momento. Assim, $p(\theta|x)$ é a *posteriori* de θ em relação a X (que já foi observado) mas é a *priori* de θ em relação a Y (que não foi observado ainda). Após observar $Y = y$ uma nova *posteriori* (relativa a $X = x\ e\ Y = y$) é obtida aplicando-se novamente o teorema de Bayes. Mas será que esta *posteriori* final depende da ordem em que as observações $x\ e\ y$ foram processadas? (CARLIN; LOUIS, 2000).

Observando-se as quantidades $x_1, x_2 \ldots, x_n$, independentes dado θ e relacionadas a θ através de $p_i(x_i|\theta)$ segue que;

$$\begin{aligned} p(\theta|x_1) &\propto l_1(\theta; x_1)p(\theta) \\ p(\theta|x_2x_1) &\propto l_2(\theta; x_2)p(\theta|x_1) \\ &\propto l_2(\theta; x_2)\, l_1(\theta; x_1)p(\theta) \\ &\quad \ldots\ldots\ldots \\ &\quad \ldots\ldots\ldots \\ &\quad \ldots\ldots\ldots \\ p(\theta|x_n, x_{n-1}, \ldots, x_1) &\propto \left[\prod_{i=1}^{n} l_i(\theta; x_i)\right] p(\theta) \\ &\propto \; l_n(\theta; x_n)\, p(\theta|x_{n-1}, \ldots, x_1) \end{aligned} \tag{11}$$

Ou seja, a ordem em que as observações são processadas pelo teorema de Bayes é irrelevante. Na verdade, elas podem até ser processadas em subgrupos.

1.7.1. PRINCÍPIO DA VEROSSIMILHANÇA

Segundo Carlin e Louis (2000), o principio da Verossimilhança postula que para fazer inferência sobre uma quantidade de interesse θ só importa aquilo que foi realmente observado e não aquilo que "poderia" ter ocorrido, mas efetivamente não ocorreu.

1.7.2. PROBABILIDADE INCONDICIONAL

"A probabilidade a priori também chamada de probabilidade incondicional, de um evento é a probabilidade atribuída a um evento na ausência de conhecimento que suporte a sua ocorrência ou ausência, isto é, a probabilidade do evento anterior a qualquer evidência" (LUGER, 2004) é simbolizada por $P(evento)$.

Para Carlin e Louis (2000), a utilização de informação a priori em inferência Bayesiana requer a especificação de uma distribuição a priori para a quantidade de interesse θ. Esta distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre θ antes da realização do experimento.

1.7.3. PROBABILIDADE CONDICIONAL

"A probabilidade a posteriori (após o fato), também chamada probabilidade condional, de um evento é a probabilidade de um evento dada alguma evidência" (LUGER, 2004) é simbolizada por $P(evento|evidência)$.

1.7.4. ESTIMAÇÃO

A distribuição a posteriori de um parâmetro θ contém toda a informação probabilística a respeito deste parâmetro e um gráfico da sua função de densidade a posteriori é a melhor descrição do processo de inferência. No entanto, segundo Carlin e Louis (2000), algumas vezes é necessário resumir a informação contida na posteriori através de alguns poucos valores numéricos. O caso mais simples é a estimação pontual de θ onde se resume a distribuição a posteriori através de um único número, θ.

1.8 SELEÇÃO DE VARIÁVEIS

Depois de feito o tratamento das variáveis contínuas e assim definido o conjunto de possíveis variáveis preditoras a ser utilizado no modelo, o próximo passo é definir quais são as variáveis mais significativas. O objetivo consiste em eliminar variáveis redundantes ou irrelevantes e como resultado, obter um modelo que aja com prudência, isto é, que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Pode ser adotada a seleção de variáveis para regressão logística conforme estudos de Hosmer e Lameshow (2000), onde existem três formas de seleção de variáveis: Forward; Backward e Stepwise.

1.9 VERIFICAÇÃO DE DEPENDÊNCIA

A separação d ou d-separação foi proposta por Pearl (1986). A idéia básica deste critério é: dois conjuntos de nós X e Y são d-separados por um conjunto de nós C se todo caminho não direcionado de um nó em X para um nó em Y é d-separado (bloqueado) por nós de C; neste caso, X e Y são condicionalmente independentes, dado C.

1.10 DISCRETIZAÇÃO DE VARIÁVEIS

A base de dados original possui variáveis discretas e contínuas. O processo de discretização consiste na transformação das variáveis contínuas em discretas. As redes bayesianas têm como premissa que todas as suas variáveis sejam categóricas, além de resultar em uma rede bayesiana mais adequada ao domínio do problema, tornam tanto o processo de aprendizagem como processo de inferência Bayesiana mais simples e eficiente, porque: O surgimento de valores discrepantes (*outliers*) normalmente afeta os resultados dos modelos, atrapalhando sua interpretação; Para usuários e especialistas, variáveis discretas são mais fáceis de entender; Com o objetivo de classificação, é suficiente estimar a probabilidade de uma observação pertencer a um determuinado intervalo.

1.11 ALGORITMOS PARA COMPUTAÇÃO BAYESIANA

Este tópico tem como objetivo apresentar os algoritmos de inferência que possa vir a ser desenvolvidos em uma continuação deste processo de modelagem.

No processo de inferência bayesiana há três tipos distintos de algoritmos de inferência segundo Castilho e Gutierrez (1997): exatos, aproximados e simbólicos. Um algoritmo de inferência denomina-se exato se as probabilidades dos nós são calculadas sem outro erro senão o de arredondamento, inerente a limitações de cálculo dos computadores. Os algoritmos aproximados utilizam distintas técnicas de simulação para obter valores aproximados das probabilidades. Em geral, estes algoritmos são utilizados em casos em que os algoritmos exatos não são aplicáveis, ou o custo computacional é elevado. Já os algoritmos simbólicos podem operar tanto com parâmetros numéricos quanto com parâmetros simbólicos, obtendo probabilidades na forma simbólica, em função dos parâmetros. Dar-se ênfase aos métodos exatos e aproximados já que apresentam bons resultados no geral.

1.11.1. MÉTODOS EXATOS

"Um método é denominado exato se realiza o cálculo das probabilidades a posteriori através de somatórios e combinações de valores, sem outro erro que não seja de arredondamento no cálculo" (CASTILHO E GUITIERREZ, 1997).

1.11.1.1. ALGORITMO DE ENUMERAÇÃO

A idéia básica do algoritmo de enumeração apresentado na equação (12) é avaliar a equação (14), sem ter que montar explicitamente a tabela de probabilidade conjunta total. Apenas, percorrem-se os nós da rede propagando as evidências e extraindo as probabilidades para que sejam feitos os somatórioas e multiplicações necessárias.

$$\sum_{i=1}^{n} P(D = d_i) = 1 \tag{12}$$

$$P(a) = \sum_{e_i \,\epsilon\, e\,(a)} P(e_i) \tag{13}$$

$$P(X|e) = \propto (X, e) = \propto \sum_{y} P(X, e, y) \tag{14}$$

A equação (12) significa que qualquer distribuição de probabilidades sobre uma única variável discreta deve somar 1. Segundo Russel e Norvig (2004), também que qualquer distribuição de probabilidade conjunta sobre qualquer conjunto de variáveis deve somar 1: isso pode ser verificado criando-se uma única megavariável cujo domínio é o produto cruzado das variáveis originais.

A equação (13) fornece um método para calcular a probabilidade de qualquer proposição, dada uma distribuição conjunta total que especifique as probabilidades de todos os eventos atômicos.

Pode-se extrair das equações (12) e (13) um procedimento geral de inferência bayesiana. Denota-se X uma variável de consulta, Eo conjunto de variáveis de evidência, e o conjunto de valores observados para E, e Y as variáveis restantes não-observadas (denominadas variáveis ocultas). Então, uma consulta $P(X|e)$ pode ser avaliada como mostra a equação (14).

A equação (14) permite responder qualquer consulta $P(X|e)$ a partir da distribuição conjunta total da rede bayesiana.

Desse modo a complexidade de espaço do algoritmo de Enumeração é linear em relação ao número de variáveis, e sua complexidade de tempo para uma rede com *n* variáveis booleanas é 2^n.

1.11.1.2. ALGORITMO DE ELIMINAÇÃO DE VARIÁVEIS

O algoritmo de Enumeração, descrito na seção anterior, pode ser substancialmente melhorado eliminando-se cálculos repetidos da esquação (14). A idéia é efetuar os cálculos apenas uma vez e guardar os resultados para uso posterior. Essa é uma forma de programação dinâmica.

Existem várias versões desse algoritmo se atemos apenas a versão de Russel e Norvig (2004). Dada uma rede bayesiana sobre um conjunto de variáveis $vars$ um conjunto de evidência E com seus respectivos valores observada em e e uma variável de consulta X_q, a computação de $P(X_q | e)$ envolve normalmente apenas um subconjunto de distribuições de probabilidade associadas à rede. Se a distribuição de probabilidades $P(X_i) \mid pa(X_i))$ é necessária para o cálculo da consulta, então diz que X_i é uma variável de requisito. O conjunto de variáveis de requisito é denotado por X_R. Variáveis de consulta X_q pertencem necessariamente a X_R, mas nem todas as variáveis E pertencem a X_R (apenas as variáveis de consulta que têm pais não-evidênciaem X_R pertencem a X_R). Dada essa notação, pode-se transformar a equação (14) na equação (15):

$$P(X|e) = P(X_q, e) = \sum_{X_R \setminus \{X_q, E\}} \left(\prod_{X_i \in X_R} P(X_i) \mid pa(X_i)) \right) \quad (15)$$

O algoritmo de eliminação de variáveis coloca as várias distribuições num vetor de distribuições. Essas distribuições são chamadas fatores. Coletam-se todos os fatores que contêm a variável X_1, retira-os do vetor de distribuições, constrói-se uma nova distribuição não normalizado $P(fil\ \ os(X_1) \mid pais(X_1),\ \ tios(X_1))$ e adiciona essa distribuição ao vetor de distribuição.

O resultado dessa operação é que X_1 foi "eliminado". Depois, sobre a variável de X_2, coletam-se todos os fatores que contêm X_2, retira-os do vetor de distribuições, multiplica as distribuições e elimina-se também X_2. O resultado dessa operação é novamente um fator, que é adicionado ao vetor de distribuições. Continua-se essa operação até se eliminar todas as variáveis possíveis. No final restará pelo menos um fator para a variável de consulta X_q. Multiplicando esses fatores juntos e normalizando o resultado, tem-se $P(X_2 | e)$.

A idéia é multiplicar os membros do vetor de fatores na seqüência dada pela ordenação. O algoritmo tenta eliminar as variáveis o mais rápido possível, para armazenar os produtos intermediários num tamanho razoável.

1.11.2. MÉTODOS APROXIMADOS

Os algoritmos considerados dentro do grupo de métodos aproximados utilizam distintas técnicas de simulação para obter valores aproximados das probabilidades. De acordo com Castilho e Gutierrez (1997) estes métodos podem ser classificados em algoritmos de simulação estocástica, métodos de simplificação de modelos e métodos baseados em busca e propagação de crença em ciclos.

A modelagem proposta aborda a simulação estocástica. Estes algoritmos também são conhecidos como algoritmos de amostragem estocástica. De acordo com Jensen (2001), a idéia principal deste método aproximado é usar o modelo da rede bayesiana para simular o fluxo do impacto ou influência da evidência sobre o resto das variáveis. Neste tipo de algoritmo, de acordo com as tabelas de probabilidade condicional da rede, gera-se um conjunto de amostras selecionadas aleatoriamente, então se realiza inferência, isto é, aproximam-se probabilidades de variáveis de "consulta" pela freqüência das suas aparições na amostra. A exatidão dos resultados vai depender do tamanho das amostras (do número de iterações que geram as amostras) e diferentemente dos métodos exatos, a estrutura da rede não é relevante no cálculo da inferência, sendo essa uma de suas vantagens principais.

1.11.2.1. ALGORITMO *FORWARD SAMPLING*

A espécie mais simples de processo de amostragem aleatória para redes bayesianas gera eventos a partir de uma rede que não tem nenhuma evidência associada a ela. A idéia é fazer a amostragem uma variável de cada vez, em ordem topológica. A distribuição de probabilidade a partir da qual se obtêm uma amostra do valor está condicionada aos valores já atribuídos aos pais da variável. A partir de observações do processo de amostragem temos que:

$$S_{PS}(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i) \mid pa(X_i) \tag{16}$$

Porque cada etapa de amostragem depende apenas dos valores dos pais. Essa expressão também é a probabilidade do evento de acordo com a representação da rede bayesianana distribuição conjunta, conforme equação (17).

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i) \mid pa(x_i) \tag{17}$$

Isto é $S_{PS}(X_1, \dots, X_n) = P(x_1, \dots, x_n)$. Este fato permite responder a "consultas" utilizando amostras.

> Em qualquer algoritmo de amostragem, as respostas são calculadas efetuando-se a contagem das amostras reais geradas. Supondo que existam *N* amostras ao todo, e seja $N\,x(_1, \ldots, x_n)$ a freqüência do evento $x_1, \ldots, x_n$. Espera-se que essa freqüência venha a convergir, no limite, para seu valor esperado de acordo com a probabilidade da amostragem (RUSSEL e NORVIG, 2004).

Consequentemente, no limite de N muito grande, espera-se que a resposta convirja para o resultado exato. Sempre que se usa uma igualdade aproximada (" ≈ ") no que se segue, quer-se indicar exatamente que a probabilidade estimada se torna exata no limite de uma amostra grande. Tal estimativa é chamada consistente.

1.11.2.2. ALGORITMO *LIKELIHOOD WEIGHTING*

O algoritmo *Likelihood Weighting* é o método de simulação estocástica mais implementado para inferência em redes bayesianas, em parte por causa da sua fácil implementação e rápido tempo de convergência comparado com o algoritmo *Forward Sampling.*

O algoritmo fixa os valores para as variáveis de evidência E e efetua a amostragem apenas das amostragens restantes X e Y. Garantindo que cada evento gerado será consistente com a evidência. Porém, nem todos os eventos são iguais. Antes de efetuar as contas na distribuição para a variável de consulta, cada evento é ponderado pelaprobabilidade de que o evento concorde com a evidência, medida pelo produto das probabilidades condicionais para cada variável de evidência, dados os seus pais. Intuitivamente, eventos em que a evidência real parece improvável devem receber menor peso.

Para entender por que a ponderação de probabilidades funciona, precisa-se examinar a distribuição de amostragem S_{WS} para o algoritmo. Continua-se com a notação em que o conjunto de variáveis de evidência E possui um conjunto de valores observados e. Denominam-se as outras variáveis de *Z*, isto é, $Z = \{X\} \cup Y$. O algoritmo realiza a amostragem de cada variável em *Z*, dados os valores de seus pais:

$$S_{WS}(z, e) = \prod_{i=1}^{l} P(z_i) \mid pa(Z_i) \qquad (18)$$

Observa-se que $pa(Z_i)$ pode incluir ao mesmo tempo variáveis de consulta e variáveis de evidência. Diferente da distribuição a priori P(z), a distribuição S_{WS} dedica alguma atenção à evidência: os valores amostrados para cada Z_i serão influenciados pela evidência entre os ancestrais de Z_i. Por outro lado S_{WS} dedica menor atenção à evidência do que a distribuição posterior verdadeira $P(z|e)$. "O peso da probabilidade *w* constitui a diferença entre as dis-

tribuições de amostragem real e desejada" (RUSSEL e NORVIG, 2004). O peso para uma dada amostra $\mathbf{x}$, composta de $\boldsymbol{z}$ e $\boldsymbol{e}$ é o produto das probabilidades para cada variável de evidência, dados seus pais:

$$w(z,e) = \prod_{i=1}^{m} P(e_i) \mid pa(E_i) \tag{19}$$

Multiplicando as equações (18) e (19), pode-se observar que a probabilidade ponderada de uma amostragem tem uma forma particularmente conveniente:

$$S_{WS}(z,e) \times w(z,e) = \prod_{i=1}^{l} P(z_i) \mid pa(Z_i) \times \prod_{i=1}^{m} P(e_i) \mid pa(E_i) = \boldsymbol{P}(\boldsymbol{y},\boldsymbol{e}) \tag{20}$$

Pois os dois produtos abrangem as variáveis de toda a rede, permitindo utilizar-se da probabilidade conjunta. Pode-se mostrar que as estimativas de ponderação de probabilidades são consistentes. Para quaisquer valores específicos x de X, a probabilidade posterior pode ser calculada como:

$$\begin{aligned} P(X|e) &= \alpha \sum_{y} N_{WS}(x,y,e) \times w(x,y,e)\ a\ partir\ do\ algoritmo \\ &= \alpha' \sum_{y} S_{WS}(x,y,e) \times w(x,y,e)\ para\ \boldsymbol{N}\ grande \\ &= \alpha' \sum_{y} P(x,y,e)\ pela\ equação\ (20) \\ &= \alpha'\, \boldsymbol{P}(x,\boldsymbol{e}) = P(x,\boldsymbol{e}) \end{aligned} \tag{21}$$

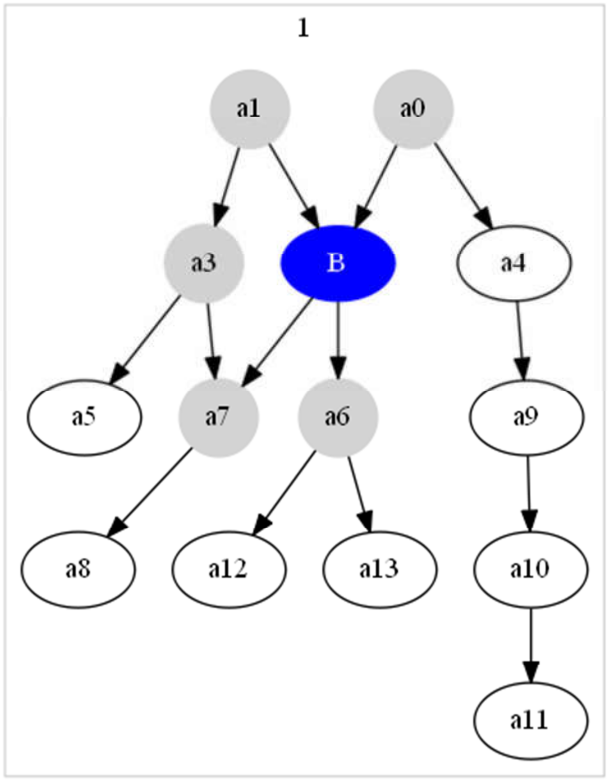
Consequentemente a ponderação de probabilidades retorna estimativas consistentes. Tendo em vista que a ponderação de probabilidade utiliza todas as amostras geradas, ela pode ser muito mais eficiente que o algoritmo apresentado na seção anterior. Entretanto ele sofrerá uma degradação de desempenho à medida que o número de variáveis de evidência aumentar. Como muitas amostras terão pesos muito baixos, consequentemente a estimativa ponderada será dominada pela minúscula fração de amostras que concordam em uma proporção maior que uma probabilidade infinitesimal com a evidência.

1.11.2.3. ALGORITMO *GIBBS SAMPLING*

Pode-se dizer que um nó é condicionalmente independente de todos os outros nós de uma rede dados seus pais, filhos e pais de seus filhos. Isto é, dada sua Cobertura de Markov.

Essa afirmação é equivalente à especificação de que um nó é independente de seus não-descendentes, dados seus pais a figura 16 ilustra essa cobertura.

Figura 2 – Cobertura de Markov. Fonte: Casella e George (1992).



Em matemática, física ou estatística, Gibbs Sampling é um método usado para gerar uma seqüência de amostras da distribuição de probabilidade conjunta das variáveis aleatórias. Segundo Casella e George (1992), o propósito dessa sequência é aproximara distribuição conjunta ou computar uma integral. O algoritmo foi nomeado depois que o físico J.W. Gibbs, em referência a analogia entre o algoritmo de amostragem e a física estatística. O algoritmo foi inventado por Geman e Geman (1984) cerca de oito décadas depois da passagem de Gibbs.

Diferentemente dos outros algoritmos de amostragem, que geram cada evento a partir do nada, o Gibbs Sampling gera cada evento fazendo uma mudança aleatória no evento precedente. Portanto, deve-se pensar que a rede se encontra em um determinado estado atual especificando um valor para uma das variáveis não de evidência X, condicionadas sobre os valores atuais das variáveis na cobertura de Markov. Então, o algoritmo vagueia ao acaso pelo espaço de estados, o espaço de atribuições completas possíveis, invertendo uma variável de cada vez, mas mantendo fixas as variáveis de evidência.

O algoritmo começa com uma configuração das variáveis consistente com as evidências, e então troca aleatoriamente o estado das outras variáveis condicionadas à sua cobertura de Markov. Depois é usada essa nova configuração gerada pra trocar os valores das outras variáveis.

Para Russel e Norvig (2004), o processo de amostragem se fundamenta em um equilíbrio dinâmico no qual a fração ao longo do prazo do tempo gasto em cada estado é exatamente proporcional à sua probabilidade posterior. Essa propriedade decorre da probabilidade de transição específica com que o processo passa de um estado para o outro, definida pela distribuição dada pela cobertura de Markov da variável cuja amostra está sendo coletada.

Seja q(x → x') a probabilidade de que o processo faça uma transição do estado x para o estado x'. Essa probabilidade define o que se denomina cadeia de Markov sobre o espaço de estados. Agora, supondo que executemos a cadeia de Markov para t etapas e seja π t(x) a probabilidade de que o sistema esteja no estado x no tempo t. De modo semelhante, seja π t+1(x') a probabilidade de o sistema se encontrar no estado x' no tempo t+1. Dado π t(x) pode-se calcular π t+1(x') efetuando o somatório da probabilidade de estar em um estado multiplicada pela probabilidade de fazer a transição para x', para todos os estados em que o sistema poderia se encontrar no tempo t:

$$\pi_{t+1}(x') = \sum_{x} \pi_t(x)\, q(x \rightarrow x') \tag{22}$$

Diz-se que a cadeia alcançou sua distribuição estacionária se $\pi_{t+1} = \pi_t$. Essa distribuição estacionária é denotada por π; e sua definição é:

$$\pi(x') = \sum_{x} \pi(x)\, q(x \rightarrow x')\ para\ todo\ \boldsymbol{x} \tag{23}$$

Sob certas suposições-padrão sobre a distribuição de probabilidade de transição q, existe exatamente uma distribuição π que satisfaz a essa equação para qualquer q dado. A equação (23) pode ser interpretada com o significado de que o "fluxo de saída" esperado a partir de cada estado é igual ao "fluxo de entrada" de todos os estados. Um modo de satisfazer a esse relacionamento ocorre se o fluxo esperado entre qualquer par de estados é o mesmo em ambos os sentidos, segundo Jensen (2001) essa é a propriedade de equilíbrio detalhado.

$$\pi(x)q(x \rightarrow x') = \pi(x')\, q(x \rightarrow x')\ \ para\ todo\ \boldsymbol{x}, \boldsymbol{x'} \tag{24}$$

Pode-se mostrar que o equilíbrio detalhado implica imutabilidade efetuando o somatório sobre x na equação (24). Tem-se que:

$$\sum_{x} \pi(x)\, q(x \to x') = \sum_{x} \pi(x')\, q(x' \to x) = \pi(x') \tag{25}$$

Onde a última etapa se segue porque se tem a garantia de que irá ocorrer uma transição a partir de x'.

Agora, pode ser demonstrado que a probabilidade de transição $q(x \to x')$ definida pela etapa de amostragem em Gibbs Sampling satisfaz à equação de equilíbrio detalhado com uma distribuição estacionária igual a $P(x|e)$.

Para tal demonstração, primeiro define-se uma cadeia de Markov no qual que se efetuam amostras condicionais de cada variável sobre os valores atuais de todas as outras variáveis, e mostra-se que isso satisfaz ao equilíbrio detalhado. Em seguida, observa-se que, para redes bayesianas, isso é equivalente a fazer a amostragem condicional sobre a cobertura de Markov da variável.

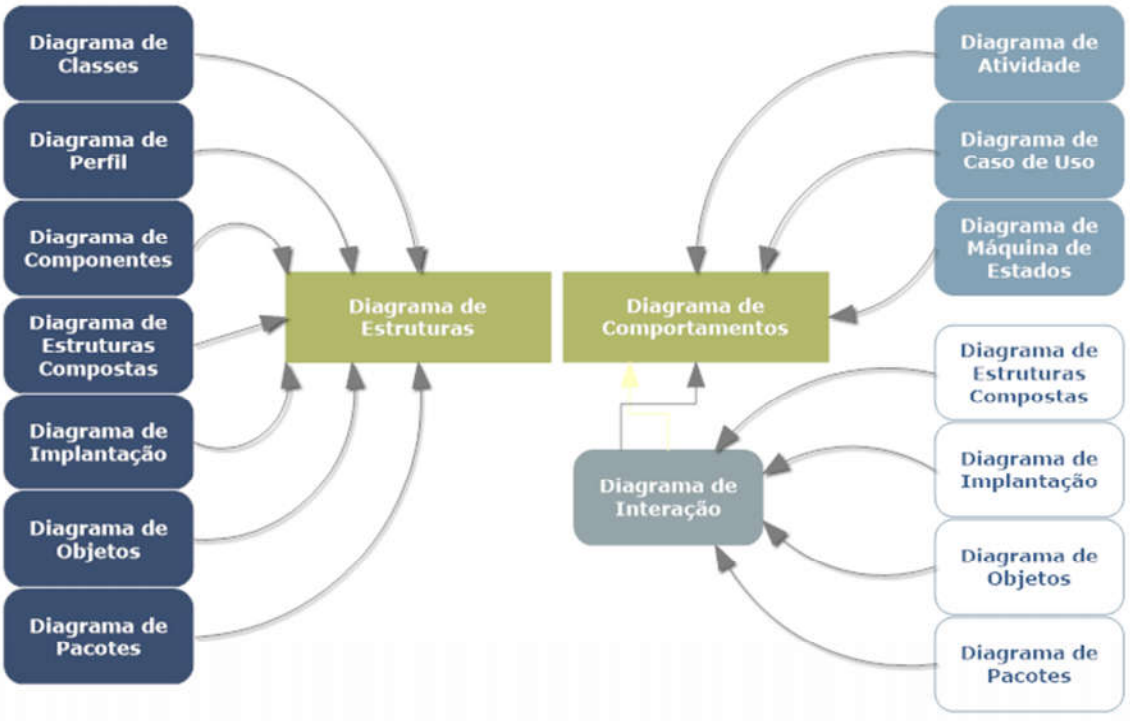
1.12 LINGUAGEM DE MODELAGEM UNIFICADA – UML

Em Engenharia de Software ou Desenvolvimento de Sistemas existem muitas metodologias para auxiliar neste processo que é a "produção de um aplicativo". Neste contexto a linguagem de modelagem unificada (do inglês, UML - *Unified Modeling Language*) é uma linguagem de modelagem que permite representar um sistema de forma padronizada ou notação de diagramas para especificar, visualizar e documentar modelos de "software", não só orientados por objetos, mas também documentar a funcionalidade de determinada aplicação. "Modelar um software tem o objetivo de tornar mais clara à visualização de suas funcionalidades, componentes, interações sejam para um desenvolvedor ou para um cliente" (OMG, 2015).

Ao invés de tentar criar seus próprios símbolos ou identificadores, utilizar algo já consistente e que é um padrão internacional como a UML, por não ser uma metodologia de desenvolvimento, o que significa que ela "não diz" o que fazer primeiro e em seguida ou como projetar algum sistema, mas auxilia a visualizar seu desenho e a comunicação entre os objetos. A UML é controlada pelo Object Management Group (OMG) e é a norma da indústria para descrever graficamente o "software" (OMG, 2015).

Os objetivos da UML são: especificação, documentação, estruturação para subvisualização e maior visualização lógica do desenvolvimento completo de um sistema de informação. Os seguintes tipos de diagramas fazem parte da UML, como descritos na figura 3:

Figura 3 - Diagramas da UML. Fonte: OMG (2015). Adaptado por Calderaro (2016).



Primariamente a UML permite que desenvolvedores visualizem os produtos de seus trabalhos em diagramas padronizados. Junto com uma notação gráfica, a UML também especifica significados, isto é, semântica. É uma notação independente de processos, embora o RUP (*Rational Unified Process*) tenha sido especificamente desenvolvido utilizando a UML.

Há uma distinção entre um modelo UML e um diagrama (ou conjunto de diagramas) de UML. O último é uma representação gráfica da informação do primeiro, mas o primeiro pode existir independentemente. O XMI (*XML Metadata Interchange*) na sua versão corrente disponibiliza troca de modelos não de diagramas.

Um sistema bem definido em sua modelagem não necessariamente precisa estar representado em todos os diagramas disponíveis na UML, mas apenas alguns são considerados essenciais para UML. Isto é, Diagrama de Casos de Uso; Diagrama de Classes; Diagrama de Objetos; Diagrama de Comunicação (antigo diagrama de colaboração); Diagrama de Sequência; Diagrama de Máquina de Estados; Diagrama de Componentes; Diagrama de Pacotes.

> "O objetivo é fornecer múltiplas visões do sistema a ser modelado, analisando-o e modelando-o sob diversos aspectos, procurando-se, assim, atingir a completitude da modelagem, permitindo que cada diagrama complemente os outros" (GUEDES, 2009).

E cada diagrama da UML analisa o sistema, ou parte dele, sob uma determinada óptica. É como se o sistema fosse modelado em camadas, sendo que alguns diagramas enfocam o

sistema de forma mais geral, apresentando uma visão externa do sistema, como é o objetivo do Diagrama de Casos de Uso, enquanto outros oferecem uma visão de uma camada mais profunda do software, apresentando um enfoque mais técnico ou ainda visualizando apenas uma característica específica do sistema ou um determinado processo. A utilização de diversos diagramas permite que falhas sejam descobertas, diminuindo a possibilidade da ocorrência de erros futuros, na figura.

Por exemplo, a construção de um edifício, percebe-se que ao projetar uma construção, esta não tem apenas uma planta, mas diversas, enfocando o projeto de construção do prédio sob diferentes formas, algumas referentes ao *layout* dos andares, outras apresentando a planta hidráulica e outras ainda abordando a planta elétrica, por exemplo. Isso torna o projeto do edifício completo, abrangendo todas as características da construção.

Quadro 2 – Diagramas da UML essenciais. Fonte: OMG (2015). Adaptado por Calderaro (2016).

UML	
Diagramas	Descrição
Diagrama de Caso de Uso	O diagrama de casos de uso é o diagrama mais geral e informal da UML, utilizando normalmente nas fases de levantamento e análise de requisitos do sistema, embora venha a ser consultado durante todo o processo de modelagem e possa servir de base para outros diagramas. Apresenta uma linguagem simples e de fácil compreensão para que os usuários possam ter uma ideia geral de como o sistema irá se comportar. “Procura identificar os atores que utilizarão de alguma forma o software, bem como os serviços, ou seja, as funcionalidades que o sistema disponibilizará aos atores, conhecidas nesse diagrama como casos de uso” (GUEDES, 2009).
Diagrama de Classe	O diagrama de classes é provavelmente o mais utilizado e é um dos mais importantes da UML. Serve de apoio para a maioria dos demais diagramas. “Como o próprio nome diz, define a estrutura das classes utilizadas pelo sistema, determinando os atributos e métodos que cada classe tem, além de estabelecer como as classes se relacionam e trocam informações entre si” (GUEDES, 2009).

Diagrama de Objetos	O diagrama de objetos está amplamente associado ao diagrama de classes. Na verdade, o diagrama de objetos é praticamente um complemento do diagrama de classes e bastante dependente deste. “O diagrama fornece uma visão dos valores armazenados pelos objetos de um diagrama de classes em um determinado momento da execução de um processo do software” (GUEDES, 2009).
Diagrama de Comunicação	O diagrama de comunicação era conhecido como de colaboração até a versão 1.5 da UML, tendo seu nome modificado para o diagrama de comunicação a partir da versão 2.0. Está amplamente associado ao diagrama de sequência: na verdade um complementa o outro. As informações mostradas no diagrama de comunicação com frequência são praticamente as mesmas apresentadas no de sequência, porém com um enfoque distinto, visto que esse diagrama não se preocupa com a temporalidade do processo, concentrando-se em como os elementos do diagrama estão vinculados e quais mensagens trocam entre si durante o processo (GUEDES, 2009).
Diagrama de Sequência	O diagrama de sequência é um diagrama comportamental que se preocupa com a ordem temporal em que as mensagens são trocadas entre os objetos envolvidos em um determinado processo. Em geral, baseia-se em um caso de uso definido pelo diagrama de mesmo nome e apoia-se no diagrama de classes para determinar os objetos das classes envolvidas em um processo. “Um diagrama de sequência costuma identificar o evento gerador do processo modelado, bem como o ator responsável por esse evento e determina como o processo deve se desenrolar e ser concluído por meio de chamada de métodos disparados por mensagens enviadas entre os objetos” (GUEDES, 2009).
Diagrama de Máquina de Estados	O diagrama de máquina de estados demonstra o comportamento de um elemento por meio de um conjunto finito de transições de estado, ou seja, uma máquina de estados. Além de poder ser utilizado para expressar o comportamento de uma parte sistema, quan-

	do é chamado de máquina de estado comportamental, também pode ser usado para expressar o protocolo de uso de parte de um sistema, quando identifica uma máquina de estado de protocolo. “Como do diagrama de sequência, o de máquina de estados pode basear-se em um caso de uso, mas também pode ser utilizado para acompanhar os estados de outros elementos, como, por exemplo, uma instância de uma classe” (GUEDES, 2009).
Diagrama de Componentes	O diagrama de componentes está amplamente associado à linguagem de programação que será utilizada para desenvolver o sistema modelado. Esse diagrama representa os componentes do sistema quando o mesmo for ser implementado em termos de módulos de código-fonte, bibliotecas, formulários, arquivos de ajuda, módulos executáveis etc. “e determina como tais componentes estarão estruturados e irão interagir para que o sistema funcione de maneira adequada” (GUEDES, 2009).
Diagrama de Pacotes	O diagrama de pacotes é um diagrama estrutural que tem por objetivos representar os subsistemas ou submódulos englobados por um sistema de forma a determinar as partes que compõem. Pode ser utilizado de maneira independente ou associado com outros diagramas. “Esse diagrama pode ser também para auxiliar a demonstrar a arquitetura de uma linguagem, como ocorre com a própria UML ou ainda para definir as camadas de um software ou um processo de desenvolvimento” (GUEDES, 2009).

Da mesma maneira, os diversos diagramas fornecidos pela UML permitem analisar o sistema em diferentes níveis, podendo enfocar a organização estrutural do sistema, o comportamento de um processo específico, a definição de um determinado algoritmo ou até mesmo as necessidades físicas do sistema para que este funcione de forma adequada.

1.13 POSTGRESQL

Quando a necessidade de guardar um grande volume de dados ou simplesmente o ato de proteger seus dados entra em ação os Sistemas Gerenciadores de Banco de Dados (SGDB), O PostgreSQL que foi pioneiro em vários conceitos que somente se tornaram disponíveis muito mais tarde em alguns sistemas de banco de dados comerciais. Possui um desenvolvimento ativo pela comunidade *open sorce* e uma arquitetura que comprovadamente ganhou forte reputação de confiabilidade, integridade de dados e conformidade a padrões. "Roda em todos os grandes sistemas operacionais, incluindo GNU/Linux, Unix (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64), e MS Windows" (POSTGRESQL, 2014).

Descendente de código fonte aberto deste código original de Berkeley, que suporta grande parte do padrão Structure Query Language (SQL) e oferece muitas funcionalidades modernas como descrito no quadro 3:

Quadro 3 – Funcionalidades. Fonte: PostgreSQL (2014). Adaptado por Calderaro (2016).

Itens	Descrição
Comandos complexos	Comandos SQL com alta disponibilidade de dados.
Chaves estrangeiras	Campo que aponta para uma chave primária de outra tabela.
Gatilhos	Procedimentos disparados antes ou após outro procedimento.
Visões	Uma tabela virtual somente leitura.
Integridade transacional	Garante que após uma falha de sistema qualquer transação de dados não fique quebrados.
Controle de simultaneidade multi-versão	Isolamento de transações mantendo a consistência dos dados.

Totalmente compatível com o acrônimo de Atomicidade, Consistência, Isolamento e Durabilidade (ACID), tem suporte completo a chaves estrangeiras, junções – Junções de Tuplas em Banco De Dados – (JOINs), visões, gatilhos e procedimentos armazenados (em múltiplas linguagens). Inclui a maior parte dos tipos de dados do ISO SQL:1999, incluindo INTEGER, NUMERIC, BOOLEAN, CHAR, VARCHAR, DATE, INTERVAL, e TIMESTAMP. "Suporta também o armazenamento de objetos binários, incluindo figuras, sons ou

vídeos. Possui interfaces nativas de programação para C/C++, Java, .Net, Perl, Python, Ruby, Tcl, ODBC, entre outros, e uma excepcional documentação" (POSTGRESQL, 2014).

Devido à sua licença liberal, o PostgreSQL pode ser utilizado, modificado e distribuído por qualquer pessoa para qualquer finalidade, seja particular, comercial ou acadêmica, livre de encargos. Este SGDB pode ser ampliado pelo usuário de muitas maneiras como, por exemplo, adicionando novos recursos listados no quadro 4:

Quadro 4 – Recursos em constante evolução. Fonte: PostgreSQL (2014). Adaptado por Calderaro (2016).

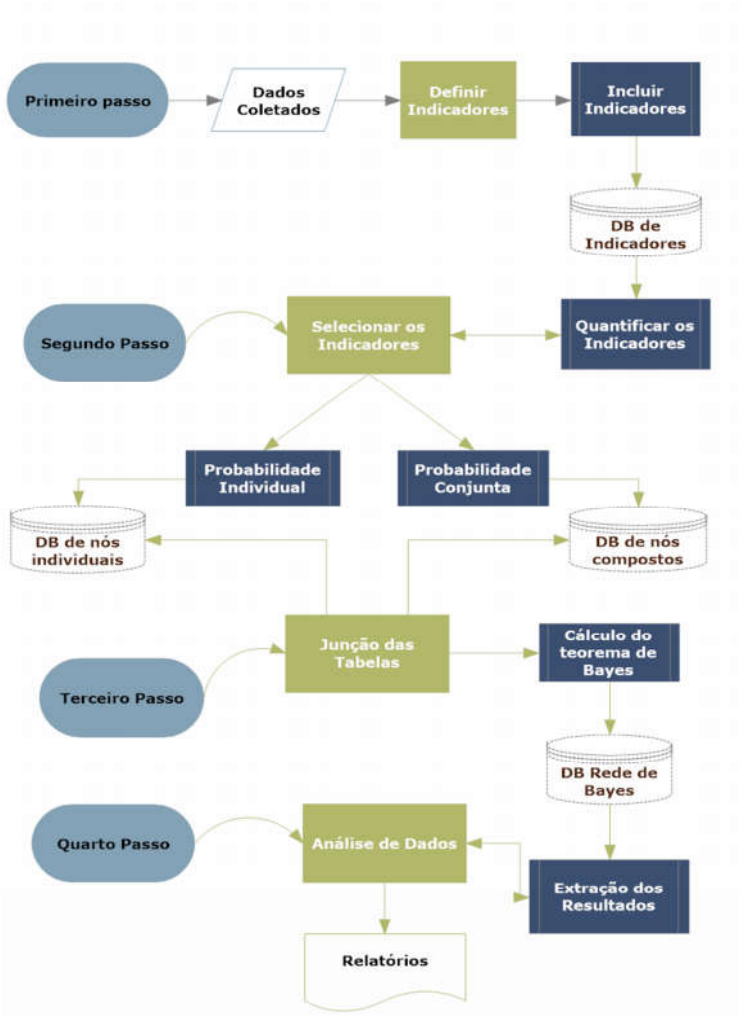
Itens	Descrição
Tipos de dado	São as definições dos dados que vão compor a tabela.
Funções	Funções matemáticas do banco.
Operadores	Operadores matemáticos do banco.
Funções de agregação	Funções de agrupamento de dados.
Métodos de índice	Organização dos dados em alguma ordem específica.
Linguagens procedurais	Linguagens para definição de funções.

CAPÍTULO 2: MATERIAIS E MÉTODOS

2.1 MÉTODOS

O presente trabalho está divido em etapas descritas na figura 1. Essas etapas são respectivamente os processos de definição, agrupamento, quantificação, seleção, junção e qualificação dos indicadores definidos.

Figura 4 – Diagrama de análise de indicadores. Fonte: Calderaro e Almeida (2014).



O diagrama de análise representa fielmente a ideia central deste trabalho, foi definido o esquema mais enxuto, evitando sempre a sobrecarga desnecessária, ou seja, a representação

gráfica de objetos que não são essenciais ao modelo a ser produzido com representação gráfica nos diagramas UML.

2.2 MÉTODO QUANTITATIVO

O método de análise quantitativa obedece a um plano preestabelecido, com o intuito de enumerar ou medir eventos, examina as relações entre as variáveis por métodos experimentais ou semi-experimentais, controlados com rigor, emprega geralmente, para a análise dos dados, instrumentos estatísticos, confirma as hipóteses da pesquisa ou descobertas por dedução, ou seja, realiza observações ou experiências, utiliza dados que representam uma população específica (amostra), a partir da qual os resultados são generalizados, e usa como instrumento para escolha de dados, questionários estruturados, elaborados com questões, testes e checklist, aplicados a partir de entrevistas individuais, apoiadas por um questionário convencional (impresso) ou eletrônico (NEVES 2006).

A Figura 5 apresenta a construção esquemática da heurística para armazenamento, análise quantitativa e seleção dos dados para armazenamento na base de conhecimento.

Figura 5 – Armazenamento, análise quantitativa e seleção dos dados para a base de conhecimento. Fonte: Calderaro e Almeida (2014).

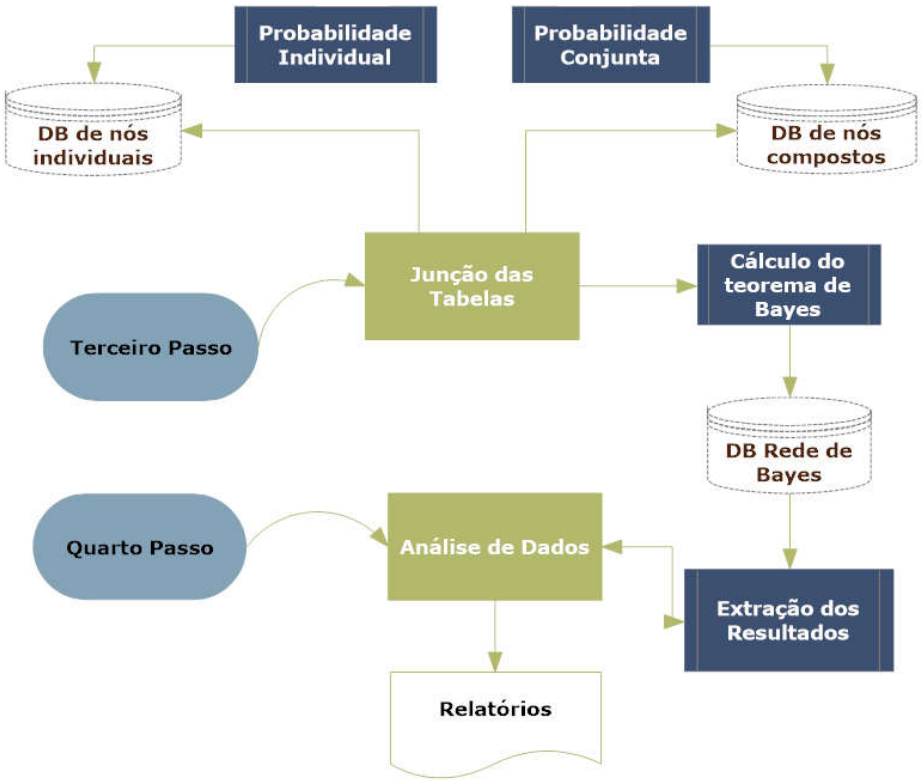
Esta parte do diagrama está representada por esquemas (Schema) no banco de dados, cada esquema possui seus próprios elementos.

2.3 MÉTODO QUALITATIVO

O método de análise qualitativa é descrito pela apresentação da descrição e análise dos dados em uma síntese narrativa, busca de significados em contextos social e culturalmente específicos, porém com a possibilidade de generalização teórica, ambiente natural como fonte de recolha de dados e investigador como instrumento principal desta atividade, tendência a ser descritivo, maior interesse pelo processo do que pelos resultados ou produtos com aquisição de dados por meio de entrevista, observação, investigação participativa, entre outros, procura da compreensão dos fenômenos, pelo investigador, a partir da perspectiva dos participantes.

E, finalmente, a utilização do enfoque indutivo na análise dos dados, ou seja, realização de generalizações de observações limitadas e específicas pelo pesquisador (NEVES 1996). Já o processo de extração, análise qualitativa e seleção de dados estão demonstradas na Figura 6.

Figura 6 – Processo de extração, análise qualitativa e seleção de dados da base de conhecimento. Fonte: Calderaro e Almeida (2014).



Esta parte do diagrama está representada por esquemas (Schema) no banco de dados, cada esquema possui seus próprios elementos.

2.4 METODOLOGIA

Esta dissertação apresenta como metodologia da pesquisa: bibliográfica e experimental, porque é aplicada no conjunto de indicadores preestabelecidos e adaptados no banco de dados a rede bayesiana baseada nos indicadores coletados no banco de dados do Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES).

A pesquisa bibliográfica é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos. Embora em quase todos os estudos seja exigido algum tipo de trabalho dessa natureza, há pesquisas desenvolvidas exclusivamente a partir de fontes bibliográficas. Boa parte dos estudos exploratórios pode ser definida como pesquisas bibliográficas. As pesquisas sobre ideologias, bem como aquelas que se propõem à análise das diversas posições acerca de um problema, também costumam ser desenvolvidas quase exclusivamente mediante fontes bibliográficas (GIL 2008).

E ainda Gil (2008) afirma que a principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente. Essa vantagem da pesquisa bibliográfica tem, no entanto, uma contrapartida que pode comprometer em muito a qualidade da pesquisa. Muitas vezes, as fontes secundárias apresentam dados coletados ou processados de forma equivocada. Assim, um trabalho fundamental nessas fontes tenderá a reproduzir ou mesmo a ampliar esses erros.

De modo geral, o experimento representa o melhor exemplo de pesquisa científica. Essencialmente, a pesquisa exeperimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-los, definir as formas de controle e de observação dos efeitos que a variável produz no objeto (GIL 2008).

Já a Pesquisa experimental consiste essencialmente em determinar um objeto de estudo, ao contrário do que faz supor a concepção popular, não precisa necessariamente ser realizada em laboratório. Pode ser desenvolvida em qualquer lugar, desde que apresente as seguintes propriedades: manipulação, controle e distribuição aleatória (GIL 2008).

O estudo de caso que irá ser apresentado conta com as seguintes características de pesquisa:

- Levantamento do material bibliográfico:

Os principais autores consultados na pesquisa científica são: (ANSAR; FLYVBJERG; BUDZIER; LUNN, 2014); (BRASIL, 2012); (FRANZIN; ALMEIDA; SOUZA, 2014); (CHARNIAK, 1991); (PEREIRA, 2008); (HABRANT, 1999); (BELLEN, 2006); (KRAMMA, 2009); (PRESCOTT-ALLEN, 1999); (MOLDAN; BILHARZ, 1997); (BAKKES, 1994);

(HARDI; BARG, 1997); (HARDI, 2000); (IDAM, 2012); (NORTH 1977); (OLIVEIRA, 2012); (AMAZONAS, 2001); (PEARCE, 2002); (BERNARDO; SMITH 1994); (GELMAN et al, 1995); (CARLIN; LOUIS, 2000); (OMG, 2015); (GUEDES, 2009); (POSTGRESQL, 2014); (SCHILDT, 1996); (CALDERARO; ALMEIDA, 2014); (PRODES 2013); (GIL 2008); (NEVES, 2006); (SANTOS, 2010); (PEREIRA E GÓES, 2013); (SANTOS, 2007), (CAVALCANTE; GÓES, 2013), (KAMPEL; CÂMARA, 2000), (DINIZ; JUNIOR; NETO; DINIZ, 2009); (PEARL, 1986); (RUSSEL; NORVIG, 2004).

- Modelagem dos indicadores:

Os indicadores utilizados para a modelagem neste estudo de caso foram extraidos do banco de dados do Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES) e foram adaptados para armazenamento no SGDB.

Para Bellen (2006) o objetivo dos indicadores é agregar e quantificar informações de modo que sua significância fique mais aparente. Os indicadores simplificam as informações sobre fenômenos complexos tentando melhorar com isso o processo de comunicação sobre eles de forma mais compreensível e quantificável. Desta forma, a quantificação é uma característica primordial para um indicador.

- Estudo aplicado da Estatística, Probabilidade e Computação Bayesiana:

Procura-se definir uma modelagem de software para a tomada de decisões, baseadas em crenças, definindo o grau de incerteza, ingenuidade heurística e um classificador probabilístico simples baseado na aplicação de teorema de Bayes. Em geral a única limitação para o número de simulações são o tempo de computação e a capacidade de armazenamento dos valores simulados. Assim, se houver qualquer suspeita de que o número de simulações é insuficiente, a abordagem mais simples consiste em simular mais valores.

- Aplicação gráfica da Computação Bayesiana com a UML:

Todo o processo computacional que irá ser desenvolvido nesta modelagem será representado graficamente atráves da UML e seus diagramas.

Os diagramas que irão ser usados na dissertação são: Diagrama de Caso de Uso, Diagrama de Sequencia, Diagrama de Comunicação.

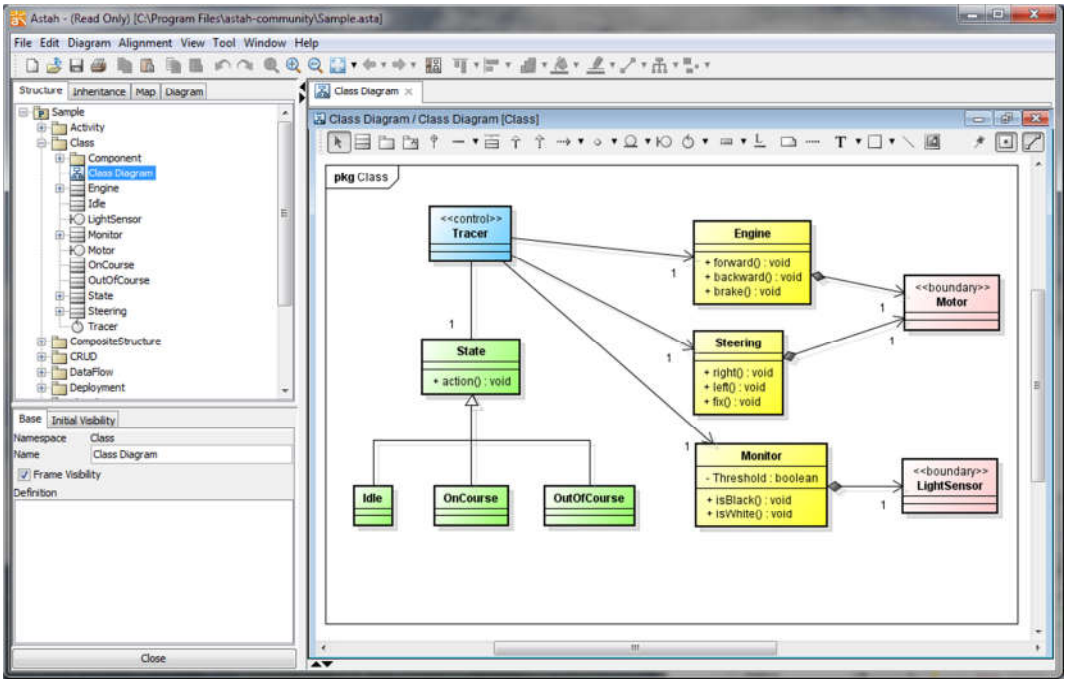
2.5 MATERIAIS

A seguir lista com todos os programas usados para o desenvolvimento da pesquisa. Foi dada a prioridade para programas livre e os pago foram comprados às referidas licenças.

2.5.1. Astah Community

Anteriormente conhecido pelo nome de *Java and UML Developers Environment* ou JUDE o Astah Comunitário é um ferramenta de extensão para a confecção dos diagramas UML, permitindo que o seus diagramas UML seja rápido e facilmente melhorado, para refinar o seu processo de desenvolvimento (ASTAH, 2014). Utilizado para a representação de todos os diagramas necessários para a análise do domínio do problema, auxiliando no desenvolvimento do software. Esta ferramenta possuí uma versão paga que não será usado neste projeto, a versão comunitária já atende amplamente o objetivo, disponível para download em http://www.astah.net.

Figura 7 – Ambiente de desenvolvimento do Astah. Fonte: http://www.astah.net.



O ambiente de desenvolvimento do Astah é intuitivo merecendo alguns poucos minutos de atenção para se adaptar.

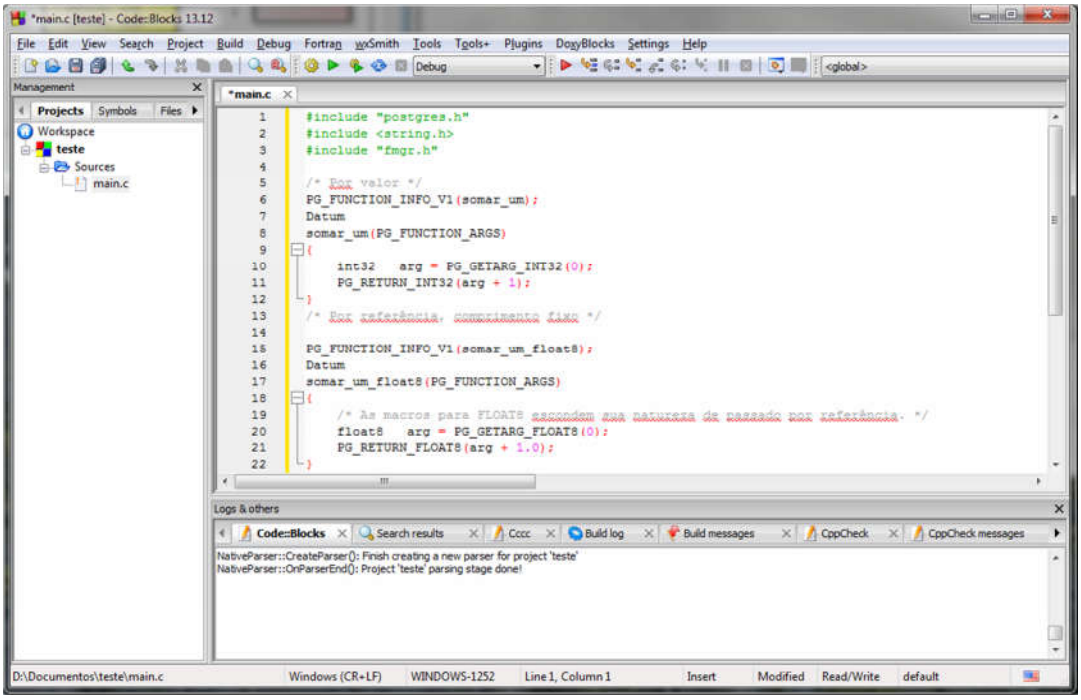
2.5.2. Code::Blocks

Construído em torno de um quadro de plug-ins o Code::Blocks pode ser estendido com plugins. Qualquer tipo de funcionalidade pode ser adicionada pela instalação / codificação de um plugin. No quadro 5 algumas vantagens do Code::Blocks

Quadro 5 – Vantagens do programa Code::Blocks Fonte: http://www.codeblocks.org.

Destaques:
• Código aberto (GPLv3), sem custos ocultos • Multi-Plataforma. É executado em Linux, Windows (utiliza wxWidgets) e Mac • Escrito C++. Sem interpretação de linguagens ou a necessidade de bibliotecas privadas. • Extensível.

Figura 8 – Ambiente de desenvolvimento Code::Blocks. Fonte: http://www.codeblocks.org.



Não muito diferente de muitas IDEs para o desenvolvimento em linguagem C, o Code::Blocks é intuitivo e ágil para confeccionar seus códigos em C.

Code::Blocks é uma IDE livre C , C ++ e Fortran construído para atender as mais exigentes necessidades de seus usuários. Ele é projetado para ser muito extensível e totalmente configurável. O Code::Blocks conta com todos os recursos necessários para o desenvolvimento nas linguagens C, C++ e Fortran, com uma aparência consistente o programa possui versões para várias plataformas disponível para download em http://www.codeblocks.org.

2.5.3. PostgreSQL Database Modeler (PgModeler)

Modelador de Banco de Dados PostgreSQL, ou simplesmente, pgModeler é uma ferramenta *open source* para bancos de dados de modelagem que funde os conceitos clássicos de diagramas entidade-relacionamento com características específicas que só PostgreSQL implementa. O pgModeler converte os modelos criados pelo usuário para código SQL e aplicá-los em *clusters* de banco de dados (PGMODELER, 2014). A figura 9 um exemplo deste ambiente.

Figura 9 – Ambiente de desenvolvimento PgModeler. Fonte: http://www.pgmodeler.com.br.

Com sua IDE diferenciada e começar como software livre, obteve sucesso junto aos desenvolvedores e possui códigos compilados pagos sem deixar de disponibilizar os códigos fontes livres para compilação disponível para download em http://www.pgmodeler.com.

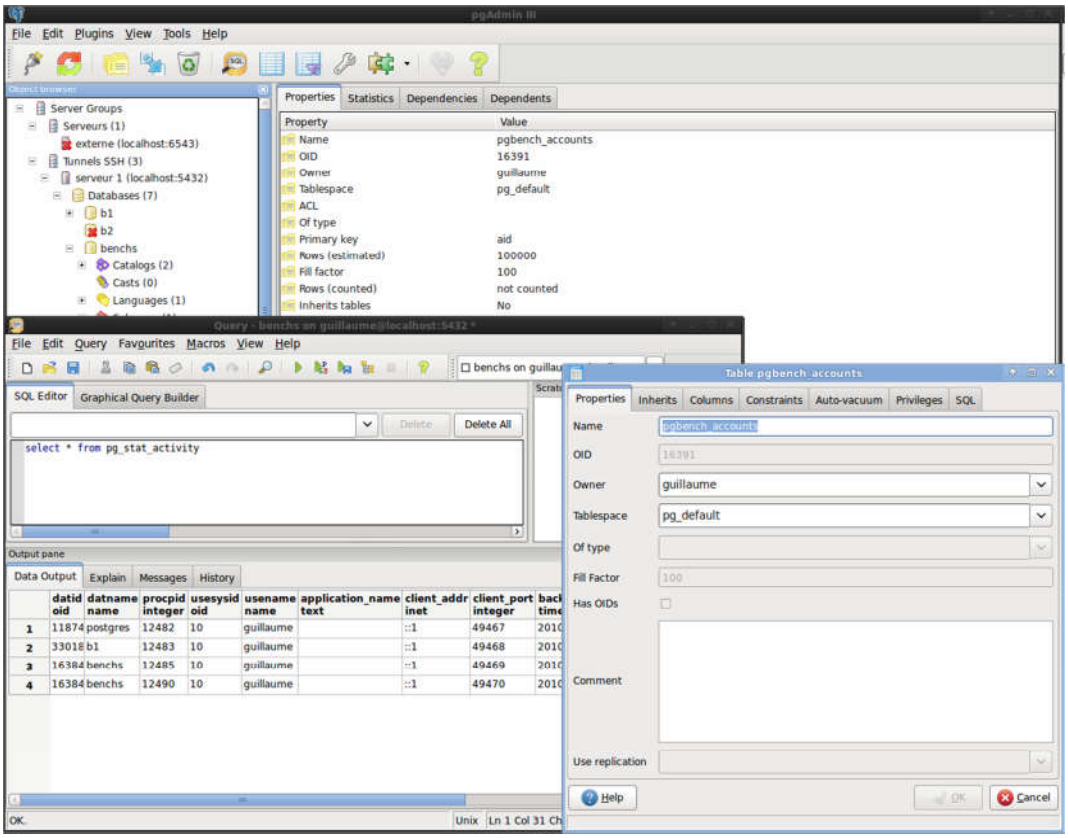
2.5.4. PgAdmin

PgAdmin é uma interface de design e gestão de banco de dados PostgreSQL, projetado para ser executado na maioria dos sistemas operacionais. O software é escrito em C ++ e usa a plataforma kit de ferramentas wxWidgets. Em cada ambiente, pgAdmin é um aplicativo nati-

vo. O aplicativo é executado em código binário, e não em uma máquina virtual, portanto, oferecendo excelente desempenho em tempo de execução.

PgAdmin é projetado para atender as necessidades de todos os usuários, desde escrever consultas SQL simples para desenvolver bases de dados complexos.

Figura 9 – Ambiente de desenvolvimento PgAdmin. Fonte: http://www.pgadmin.org.



Programa usado para gerenciar o servidor de Banco de Dados, este é o programa que vem junto da instalação do PostgreSQL. O PgAdmin é um software completo mas escaço em recursos gráficos intuitivos para facilitar o gerenciamento do PostgreSQL.

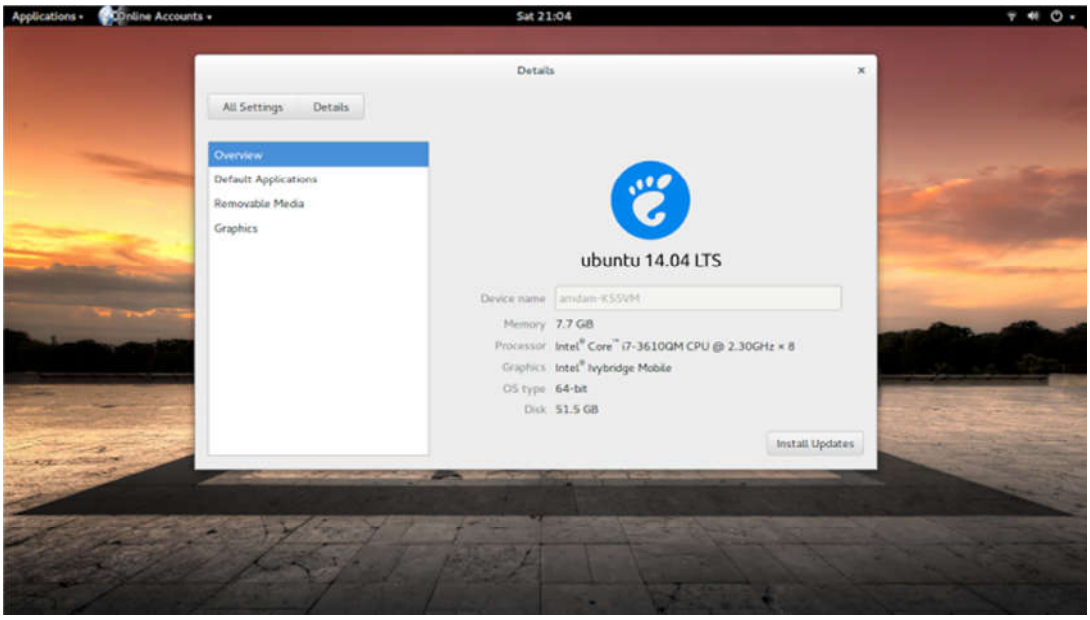
2.5.5. Linux Ubuntu 14.04 LTS

Ubuntu é um sistema operacional baseado em Linux desenvolvido pela comunidade de código livre é perfeito para notebooks, desktops e servidores. Ele contém diversos aplicativos. Ubuntu (2015).

A Canonical foi criada juntamente com o Ubuntu para ajudá-lo a chegar a um mercado mais amplo. Hoje é a instituição mantedora do Ubuntu ajudando governos e empresas de todo

o mundo com as migrações, a gestão e suporte para suas implantações do Ubuntu, Canonical (2015).

Figura 10 – Área de trabalho Ubuntu GNOME. Fonte: <https://wiki.ubuntu.com/Ubuntu GNOME/GetUbuntuGNOME



O Ubuntu possui a melhor infraestrutura de tradução e acessibilidade que a comunidade do software livre tem a oferecer, tornando o Ubuntu usável pelo máximo pessoas possíveis, Ubuntu (2015).

CAPÍTULO 3: INDICADORES E REDES BAYESIANAS

3.1 VARIÁVEIS E INDICADORES

Quando assunto é modelagem a primeira impressão é a sua similaridade com o conceito de arte, ou seja, a atividade humana ligada a manifestações de ordem estética feita por artistas a partir de percepção, emoções e idéias. O desenvolvimento de sistemas é uma arte, é a interpretação de um conceito previamente delineado para ser a solução do escopo de um problema, mas para isso existem etapas que devem ser seguidas para a conclusão deste objetivo.

Quando trata-se da modelagem de software o ponto de partida são os pré-requisitos, motivos, que levaram a ocorrência do fato observado. No âmbito desta dissertação estes pré-requisitos são as variáveis que influenciam no desmatamento. A análise bibliográfica realizada apontou a necessidade de se adicionar algumas variáveis às bases de dados já existentes, buscando relacionar desmatamento com desempenho sócio-econômico. De acordo com Pereira e Góes (2013), Neves (2006), Santos (2007), Cavalcante (2013), Kampel e Câmare (2000) e Diniz; Junior; Neto; Diniz (2009) é demonstrado no quadro 6 algumas variáveis consideradas para o desmatamento na Amazônia e na Mata Atlântica.

Quadro 6 – Lista de métodos e as variáveis que influenciam no desmatamento segundo os respectivos autores.

Autores	Variáveis	Métodos
(PEREIRA; GÓES, 2013).	Variáves macroeconômicas e do desmatamento: 1. Consumo mais gastos do governo 2. Gastos do governo 3. Consumo 4. PIB 5. FBKF 6. Exportações 7. Desmatamento	O modelo proposto neste trabalho segue os trabalhos de Hofler e Mikhail (2002) e John e Pecchenino (1994), no sentido de apresentar o conceito de qualidade ambiental a construções macroeconômicas dinâmicas e estocásticas. O ponto mais importante do modelo é a introdução da forma mais simples possível de *trade-off* entre a perda de bem-estar gerada pelo desmatamento e o ganho de produto

		agregado resultante do uso da madeira na função de produção. Mais produção significa mais consumo de bens. Então, o *trade-off* clássico entre consumo e lazer agora tem um componente a mais, que é a floresta.
(NEVES, 2006).	Variável dependente: O desmatamento. Variáveis das seguintes fontes: 1. Censos Agropecuários a. Pessoal Ocupado nos estabelecimentos agropecuários b. Utilização das terras para lavoura c. Utilização das terras para pastagens d. Efetivo de Bovinos e. Trator f. Lenha g. Madeira em toras 2. Evolução dos Remanescentes Florestais de Mata Atlântica a. Remanescentes de Mata Atlântica 3. IPEA Data a. Pib municipal b. Pib municipal industrial c. Pib municipal de serviços d. Pib municipal agrícola e. Indice de Desenvolvimento Humano (IDH) f. Custo de transporte da sede do municipio até a capital do estado.	O trabalho utiliza dois ferramentais econométricos, a correlação e o modelo de regressão linear múltipla, para identificar os tipos de relações que as variáveis possuem. A hipótese adotada é de que a conversão de áreas de florestas para uso agropecuário, ou seja, pastagens e lavouras é um fator determinante do desmatamento na região de Mata Atlântica. Assim, busca-se verificar se essa hipótese se confirma e se existem outros fatores de pressão do desmatamento na biota atlântica.
(SANTOS, 2007).	Variáveis espaciais 1. País 2. Estado 3. Município 4. Setor censitário Variáveis sintetizadas 1. Dados populacionais 2. Atividades agrícolas 3. Atividades pecuárias 4. Atividades do setor madeireiro Busca por dados oriundos da im-	A partir da lista de variáveis obtida de outros trabalhos, ficaram definidos como fatores passíveis de tabulação para serem analisados espacialmente e econometricamente os afeitos às atividades de pecuária e agricultura, setor madeireiro e dados populacionais. Eles estão presentes na quase totalidade das análises econométricas, apresentando sempre grande pertinência para explicar/prever as taxas

	plementação políticas de combate e prevenção do desmatamento não obteve êxito pois não existiam dados sistematizados sobre as ações governamentais.	de desmatamento.
(CAVALCANTE; GÓES, 2013)	Dados quantitativos de órgãos oficiais do Governo brasileiro (IBGE e INPE/PRODES)	A presente pesquisa adotou o método hipotético-dedutivo. O corte espacial sobre o qual se constituiu a base de análise deste trabalho foi baseado na perspectiva microrregional de Rondônia
(KAMPEL; CÂMARA, 2000)	Foram utilizados dados de desflorestamento da Amazônia Legal, provenientes do Projeto de Desmatamento - PRODES do INPE (1999), na forma de taxa de desmatamento (km2/ano) sobre a malha municipal do IBGE de 1994.	Análise exploratória da dependência espacial das variáveis. Verificação de autocorrelação espacial das variáveis. Verificação de associações espaciais. Regressão espacial entre variáveis.
(DINIZ; JUNIOR; NETO; DINIZ, 2009)	Variáveis de Desmatamento 1. Total de hectares desmatado por município Variáveis do setor agropecuário 1. Rebanho bovino 2. Densidade bovina 3. Culturas permanentes 4. Cultura temporária 5. Área ocupada Variáveis Socioeconômicas 1. PIB per capita 2. Educação de adultos 3. Matrícula 4. Crédito agrícola 5. População 6. Densidade demográfica	A metodologia a ser empregada se baseia em modelos dinâmicos para dados em painel, desenvolvidos por Holtz-Eakin et al. (1988) e Arellano e Bond (1991), que desenvolveram um teste de causalidade baseado no artigo seminal de Granger (1969).

Existem inúmeros trabalhos relacionados ao desmatamento que relacionam sempre as mesmas variáveis, logo na modelagem de uma Rede Bayesiana adotaremos estas variáveis extraídas de fontes como Instituto de Pesquisa Econômica Aplicada (IPEA), o Instituto Brasi-

leiro de Geografia e Estatística (IBGE) e o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), com base na literatura pertinente.

3.2 REDE BAYESIANA

Ao iniciar à estrutura de uma rede bayesiana, a principal preocupação que se deve ter é com a representação das dependências e independências condicionais. Como um grafo é utilizado, apenas as variáveis ligadas por arcos direcionados manifestam relações de dependência.

De acordo com Pearl (1986), pode-se definir um procedimento para a construção de uma rede bayesiana. Assim, dada a distribuição conjunta $P(X_1, X_2, \ldots, X_n)$ e uma determinada ordenação das variáveis, inicia-se a construção do grafo escolhendo o nó raiz X_1 e especificando sua probabilidade a priori $P(X_1)$. A seguir, acrescenta-se ao grafo o nó X_2. Se X_2 for dependente de X_1, deve-se inserir um arco direcionado com ponto inicial X_1 e ponto final X_2; feito isso, o arco deve ser quantificado com $P(X_2|X_1)$. Caso X_2 seja independente de X_1, deve-se atribuir ao nó X_2 a probabilidade a priori $P(X_2)$ e deixar os dois nós desconectados. Repetindo a operação para as demais variáveis, obtém-se a rede.

Uma vez que se define corretamente a ordem de inserção na rede, ao inserir um nó X_i sabe-se que seus pais já foram inseridos e que são os nós em $X_1, \ldots, X_{i-1}$ que influenciam diretamente X_i. O procedimento geral para a construção de uma rede bayesiana, definido por Russel e Norvig (2004) no quadro 7.

Quadro 7 – Procedimento para construção de uma rede bayesiana. Fonte: Russel e Norvig (2004).

1. Escolha o conjunto de variáveis relevantes $X_i$que descrevam o domínio;
2. Defina uma ordenação para as variáveis;
3. Enquanto restarem variáveis no conjunto:
    a. Selecione uma variável X_i e adicione um nó para ela à rede;
    b. Defina os pais de $X_i(pa(X_i))$ com algum conjunto mínimo de nós que já estão na rede, tal que a propriedade de independência condicional seja satisfeita;
    c. Defina a tabela de probabilidades condicionais de X_i.

Uma rede bayesiana representa adequadamente um domínio se cada nó de sua estrutura é independente daqueles que o precedem na ordenação dos nós, dados os nós pais. Assim, é fundamental especificar corretamente quem são os pais de cada nó, a fim de construir a estrutura correta de uma rede.

Este procedimento permite notar que todo arco com ponto final X_i tem como ponto inicial algum nó que foi inserido na rede antes deX_i. Logo, tal método de construção garante a obtenção de uma rede acíclica.

"As redes bayesianas também são livres de valores probabilísticos redundantes, o que exclui qualquer possibilidade do especialista do domínio definir uma rede que infringe os axiomas da Teoria da Probabilidade" (RUSSEL E NORVIG, 2004).

Mesmo em um domínio localmente estruturado, construir uma rede bayesiana não é uma tarefa trivial. A estrutura da rede precisa representar, sem qualquer tipo de falha, todos os agrupamentos locais de nós. A ordem correta para adicionar nós à rede é colocar primeiro as causas que não são influenciadas, depois os nós que elas influenciam, e assim sucessivamente, até chegar aos nós que não exercem influência causal direta sobre os demais. Para esse problema são implementados algoritmos de aprendizado de estrutura, que "aprendem" diretamente de uma base de dados a topologia da rede.

Após obter a topologia de uma rede bayesiana, é necessário especificar a tabela de probabilidades condicionais para cada nó da rede. Em uma tabela deste tipo, cada linha contém as probabilidades condicionais de um estado do nó para todos os casos condicionantes. Chama-se de caso condicionante uma combinação possível de valores para os nós pais.

Mesmo que um nó possua poucos pais, sua tabela de probabilidades condicionais ainda requer uma grande quantidade de números. Geralmente, preencher completamente uma tabela deste tipo é uma tarefa demorada. Para solucionar este problema, é possível programar o aprendizado automático dos parâmetros numéricos de uma rede bayesiana, desde que se conheça a sua estrutura.

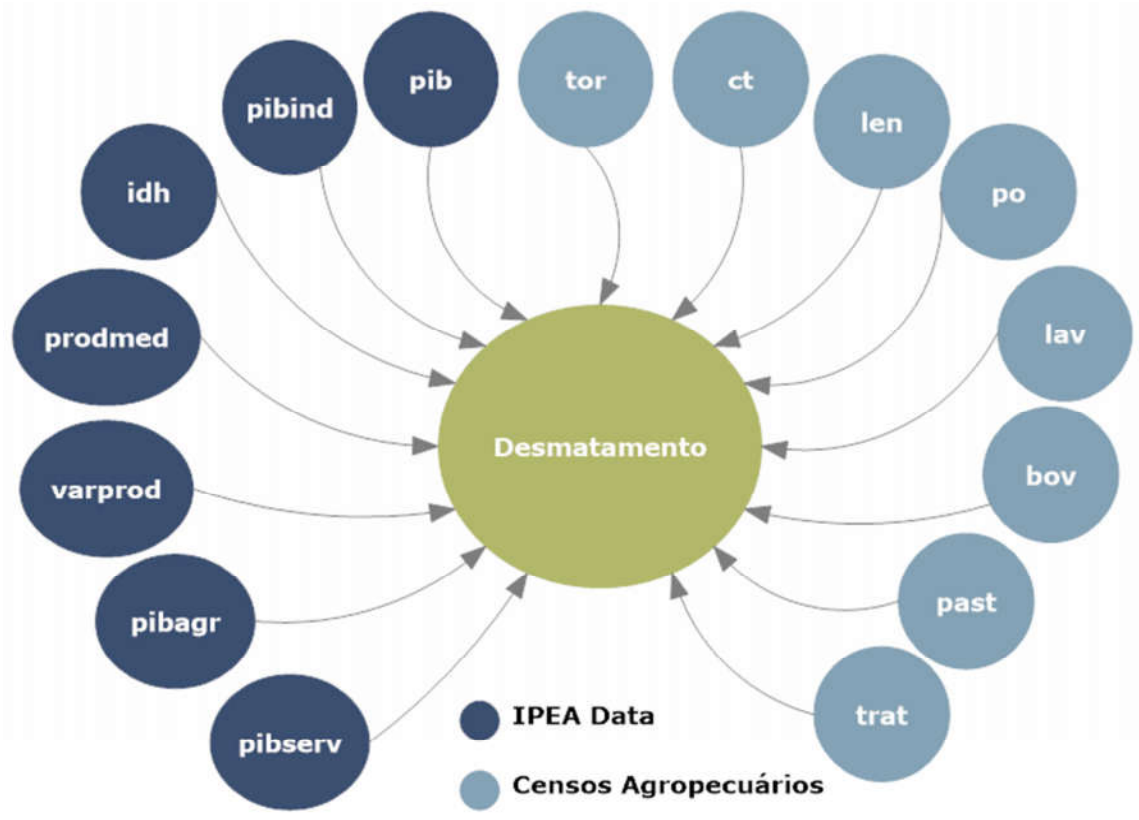
Com base nas variáveis apresentadas no quadro 06, nos referidos estudos, com a análise econométrica aplicada e tomando-se como variável dependente o desmatamento buscou-se verificar quais destas variáveis apresentam uma relação forte e significativa com o desmatamento.

Dessa forma, na investigação das causas do processo de desmatamento recente da Amazônia e a correlação entre desflorestamento e pecuária bovina no estado de Rondônia, a análise do coeficiente de correlação é uma abordagem importante e necessária para verificar se existe alguma relação entre as variáveis analisadas nos estudos descritos no quadro 6. Levan-

tamento aproveitado na construção do modelo baseado em crenças proposto neste estudo. A hipótese adotada é de que a conversão de áreas de florestas para uso agropecuário, ou seja, pastagens e lavouras é um fator determinante do desmatamento entre outras variáveis que podem ser acrescentadas. O modelo proposto representado por grafo é o mostrado na figura 11.

As variáveis que compõem a rede de acordo com as abreviações adotadas são: custo de transporte (ct), madeira em tora (tor), lenha (len), trator (trat), área de pastagem (past), rebanho bovino (bov), lavoura (lav), pessoal ocupado na área rural (po) índice de desenvolvimento humano municipal (idh), produtividade agrícola média (prodmed), variação da produtividade agrícola (varprod), PIB municipal agrícola (pibagr), PIB municipal de serviços (pibserv), PIB municipal industrial (pibind), PIB municipal (pib).

Figura 11 – Modelo proposto e sua representação em uma grafo acíclico direcionado com conexão convergente.



Onde o desmatamento não é conhecido então seus pais são independentes; caso contrário, existe correlação entre seus pais. As distribuições de probabilidades conjuntas para as duas redes são:

Censos agropecuários:

$$P(ct, tor, len, trat, past, bov, lav, po, Desmatamento) = P(ct)\ \ P(tor)\ \ P(len)\ \ P(trat)\ \ P(past)\ \ P(bov)\ \ P(lav)\ \ P(po)\ \ P(Desmatamento|ct, tor, len, trat, past, bov, lav, po)$$

IPEA Data:

$$P(id\ \ , prodmed, varprod, pibagr, pibserv, pibind, pib, Desmatamento) = P(id\ \ )\ \ P(prodmed)\ \ P(varprod)\ \ P(pibagr)\ \ P(pibserv)\ \ P(pibind)\ \ P(pib)\ \ P(Desmatamento|id\ \ , prodmed, varprod, pibagr, pibserv, pibind, pib)$$

Em diversas investigações deseja-se avaliar a relação entre duas medidas quantitativas. Três propósitos principais de tais investigações podem ser: (i) para verificar se os valores estão associados; (ii) para predizer o valor de uma variável a partir de um valor conhecido da outra; e, (iii) para descrever a relação entre variáveis. O grau de associação linear entre duas variáveis é avaliado usando correlação. Enquanto na análise de regressão as variáveis dependentes e explicativas são tratadas de forma assimétrica, na correlação quaisquer duas variáveis são tratadas simetricamente, não havendo distinção entre dependente e explicativa.

Apesar de todos os aspectos positivos associados ao formalismo de redes bayesianas, deve-se conhecer bem o domínio antes de utilizá-lo, isto porque se no domínio os relacionamentos entre as variáveis não são do tipo causal, então não se pode garantir que uma rede bayesiana é a estrutura mais apropriada para a representação das relações de dependência entre as variáveis.

O modelo na figura 11 ainda não define a relação de dependência entre as variáveis caso exista, isto ficaria a critério das relações causais que possam ser atribuídas a elas, logo o que foi proposto sofrerá alterações justamente para definir graficamente estas dependências.

Exemplo: O índice de desenvolvimento humano (idh) pode ser diretamente associado à pastagem (past) e o rebanho de bovino (bov), criando uma relação de dependência entre elas, transformando o “idh” em um nó filho das variáveis “past” e “bov”. Este tipo de associação pode ocorrer entre as demais variáveis.

CAPÍTULO 4: RESULTADOS E DISCUSSÕES

4.1 MODELAGEM DE SOFTWARE PARA TOMADA DE DECISÃO – HEURÍSTICA E COMPUTAÇÃO BAYESIANAS COM ANÁLISE UML

Segundo Neapolitan (2004), o conceito das redes bayesianas, originalmente, foi desenvolvido supondo-se uma dependência de especialistas humanos para a definição do grafo, ou seja, da estrutura ou topologia da rede e para a estimação das probabilidades condicionais, mas elas podem ser construídas tanto a partir do conhecimento de especialistas humanos quanto a partir de base de dados, com a utilização de algoritmos de aprendizado bayesianos.

A tarefa básica de um sistema de redes bayesianas é computar a distribuição da probabilidade condicional para um conjunto de variáveis de consulta, dado os valores de um conjunto de variáveis de evidência, ou seja, computar a $P(variável\_consulta|variável\_evidência)$;

Essa tarefa é denominada inferência bayesiana e permite responder a uma série de "consultas" sobre um domínio de dados. Por exemplo, no desmatamento, a principal tarefa consiste em obter um diagnóstico para uma determinada área desmatada apresentando certos fatores (evidências). Esta tarefa consiste em atualizar as probabilidades das variáveis em função das evidências.

No caso do diagnóstico para o desmatamento, tenta-se conhecer as probabilidades para o desmatamento em determinada região, de cada uma das possíveis causas das variáveis observadas. Essas são probabilidades a posteriori.

A definição dos indicadores é fundamental para a compreensão do escopo do problema. A topologia da rede bayesiana será montada com base nas variáveis. A construção da rede bayesiana exige que certos cuidados sejam tomados de forma a permitir que a tabela de conjunção de probabilidades resultantes seja uma boa representação do problema.

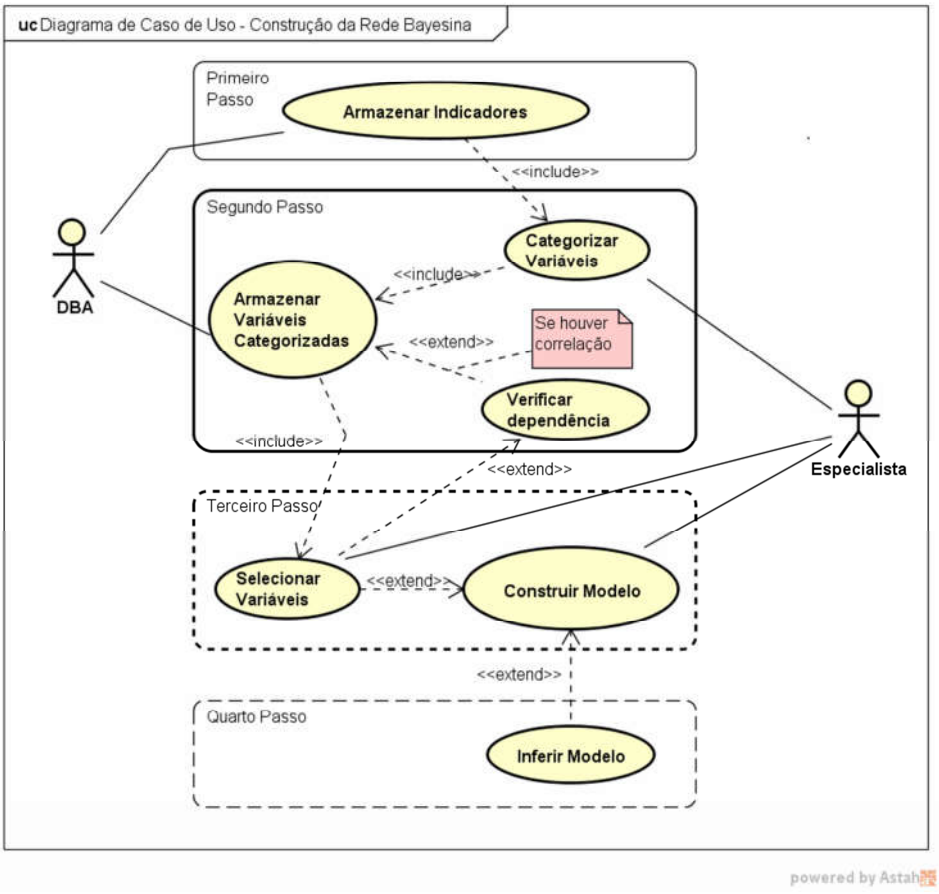
O diagrama de casos de uso auxilia na compreensão dos requisitos do sistema ajudando a especificar, visualizar e documentar as características, funções e serviços do sistema desejados pelo usuário.

Este diagrama tenta identificar os usuários que irão interagir com o sistema, quais papéis irão assumir e quais funções um usuário específico poderá requisitar. O diagrama de caso de uso tem por objetivo apresentar uma visão externa global das funcionalidades, importância

e atividade que cada ator irá desempenhar sem se preocupar como tais funcionalidades serão implementadas.

Descrevendo o diagrama de caso de uso proposto, que representa o processo de análise dos dados, na figura 12. No primeiro passo, temos o armazenamento dos indicadores que são retirados dos levantamentos apresentados no quadro 6 visto no capítulo 3, estes indicadores passam por discretização, no segundo passo, para serem categorizados de forma binária e são armazenados em outro esquema do banco de dados, é lhes conferido a dependência entre si para uma nova categorização antes de seguir para o terceiro passo, onde há uma nova seleção das variáveis para a construção do modelo bayesiano e posterior inferência realizada no quarto e último passo.

Figura 12 – Diagrama de Caso de Uso. Fonte: Calderaro (2016).

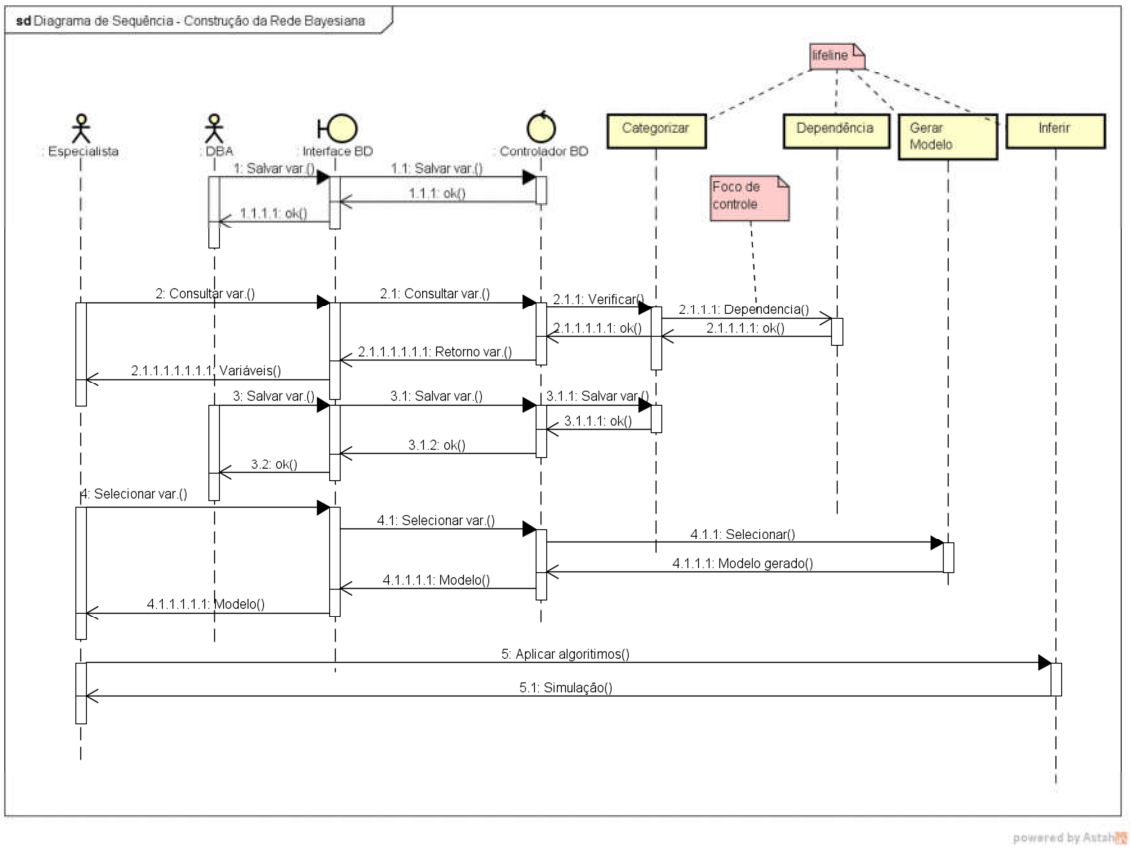


No diagrama está apresentada a associação de inclusão (*include*) que é usada quando existe um cenário, situação ou rotina comum a mais de um caso de uso. Os relacionamentos de inclusão indicam uma obrigatoriedade, ou seja, quando um determinado caso de uso tem um relacionamento de inclusão com outro, a execução do primeiro obriga também a execução

do segundo. Um relacionamento de inclusão pode ser comparado à chamada de uma subrotina ou função, artifício bastante utilizado na maioria das linguagens de programação.

O diagrama comportamental de sequência procura demonstar à sequência de eventos que ocorrem em um determinado processo, identificando quais mensagens devem ser disparadas entre os elementos envolvidos e em que ordem. A figura 13 descreve a cadência dos eventos no processo de análise de dados. Os dois agentes envolvidos no processo: especialista e DBA trocam mensagens entre si demostrando a temporalidade que é peculiar deste diagrama e do processo em si. A linha de vida representa o tempo em que um objeto (*lifeline*) existe durante um processo.

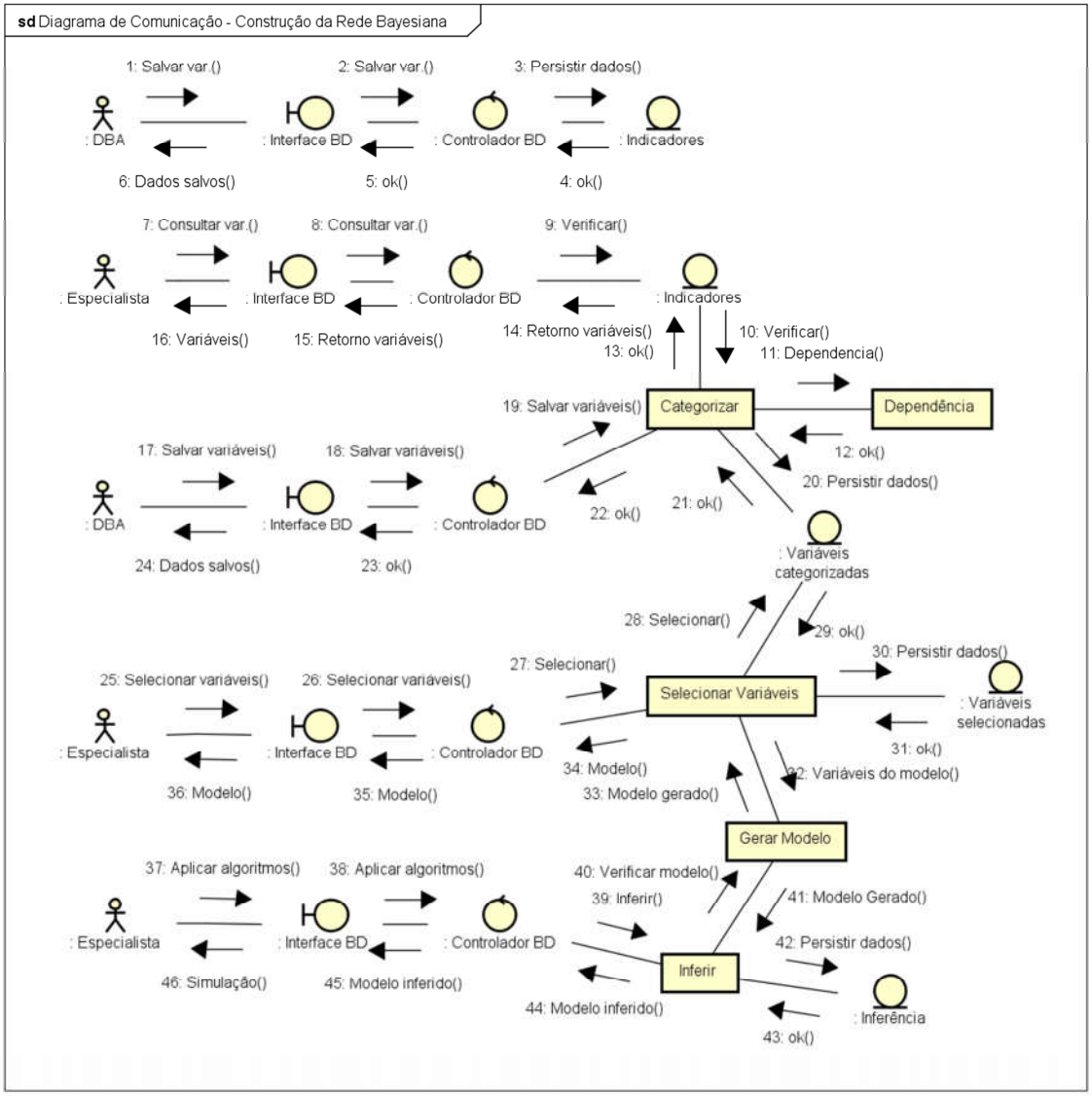
Figura 13 – Diagrama de Sequência. Fonte: Calderaro (2016).



O diagrama de sequência apresentado na figura 13, baseia-se no diagrama de casos de uso, este é um diagrama comportamental que procura determinar a sequência de eventos que ocorrem em um determinado processo, identificando quais mensagens devem ser disparadas entre elementos envolvidos e em ordem. Neste diagrama fica evidente a troca de mensagens ou estímulos para demonstrar a ocorrência de eventos, que normalmente forçam a chamada de um método em algum dos objetos envolvidos.

Da mesma forma que no diagrama de sequência, um diagrama de comunicação enfoca um processo normalmente baseado em um caso de uso. Na figura 14 temos um diagrama de comunicação.

Figura 14 – Diagrama de Comunicação. Fonte: Calderaro (2016).



As informações mostradas no diagrama de comunicação são com frequência, praticamente as mesmas apresentadas nos diagramas de sequência, porém com um enfoque diferente, visto que esses diagramas não se preocupam com a temporalidade do processo, concentrando-se em como os elementos do diagrama estão vinculados e quais mensagens trocam entre si durante um processo.

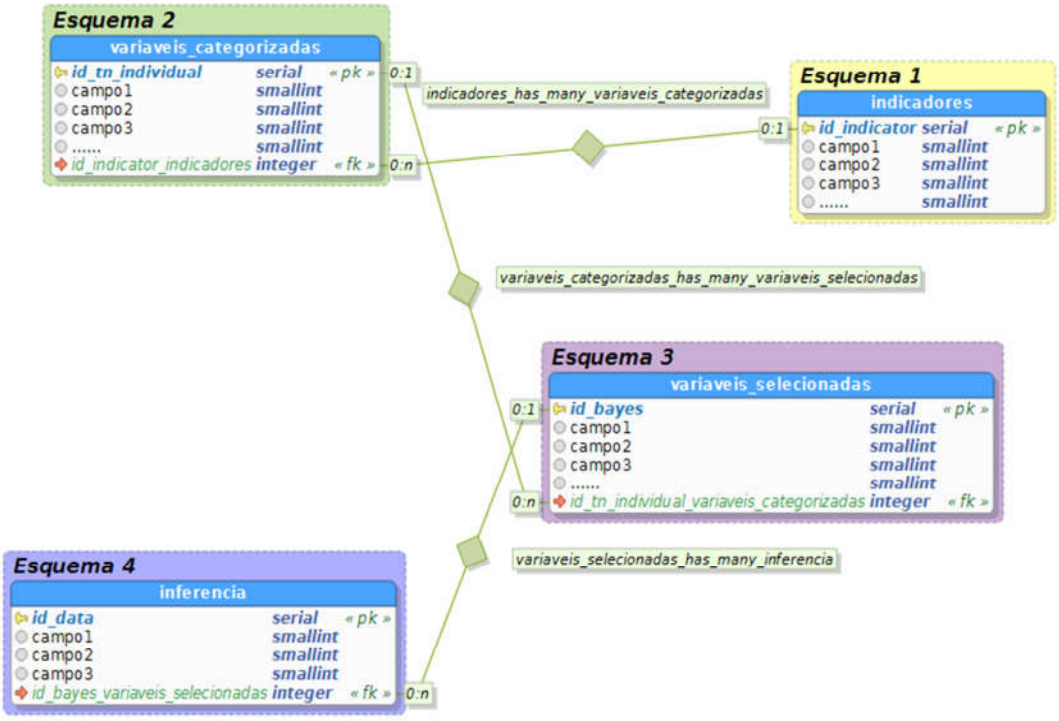
Foram acrescentadas nesse diagrama as entidades envolvidas no processo de armazenamento, seleção, categorização e inferência de dados. Observe que os agentes envolvidos sempre estão trocando mensagens com a interface do banco de dados, realizando o tratamento dos indicadores. O agente DBA se encarrega de armazenar os indicadores e suas variáveis durante os quatro passos apresentados na figura 12.

Fica a critério do especialista a consulta, seleção e aplicabilidade dos algoritmos que irão inferir o modelo proposto construído aos olhos da rede bayesiana.

4.2 ESTRUTURA DO BANCO DE DADOS PARA COMPUTAÇÃO BAYESIANA

A estrutura necessária para armazenar as variáveis é dinâmica uma vez que os indicadores e suas respectivas variáveis possuem tamanho aleatório, logo o armazenamento inicial fica a critério do administrador de banco de dados (DBA), logo abaixo na figura 15 tem-se uma eventual estrutura do banco de dados.

Figura 15 – Estrutura do banco de dados. Fonte: Calderaro (2016).



Nos relacionamentos demonstrados entre as tabelas, estes são necessários para manter a integridade dos dados. Outros relacionamentos e tabelas podem ser necessários, mas essas

podem ser criadas dinamicamente durante o processo, apenas temporariamente, no retorno de uma consulta.

Na figura 15 cada esquema representa um banco de dados separado, mas por uma questão de didática e desempenho, foi adotado que todos os esquemas estejam na mesma base de dados, eles representam as restrições importantes para a segurança com menos tempo de resposta, caso fosse necessário, estes esquemas ficam em base de dados distintas o que deve aumentar o tempo de processamento.

Portanto, um eventual contratempo é o levantamento dos indicadores e o tratamento que cada variável possa ser condicionada no decorrer do experimento da modelagem e as respectivas relações de dependência entre elas, isto define quais serão inseridas no modelo ou não através de suas correlações, ou seja, estas variáveis tem que ser incondicionais, assim como o desmatamento deverá ser dependente destas variáveis que influênciam no modelo de rede de crença proposto.

Um fator preponderante a ser identificado como risco é a incapacidade do modelo não identificar todos os pontos necessários para o desenvolvimento da aplicação ocorrendo falta de dependência em mensurar o grau de importância dos indicadores e suas variáveis, o que tornaria inviável o seu uso como elemento na tomada de decisão, que pode ser inferior ao esperado.

CONSIDERAÇÕES FINAIS

O trabalho apresentado contrubui para um pressuposto para futuras abordagens, uma delas é o desenvolvimento dos algoritmos propostos pela modelagem em um primeiro cenário, isto porque toda a lógica computacional envolve a aplicabilidade destes algoritmos.

No segundo cenário para trabalhos futuros, será desenvolvida a estrutura do banco de dados aliados à consulta SQL que deverá ser responsável pela interação entre a teoria de conjuntos e probabilidade, em um terceiro e provável último cenário, fica sob a responsabilidade da inferência Bayesiana aplicada a base de dados, toda a heurística para a extração do conhecimento.

REFERÊNCIAS

AFONSO, Rodrigo Alvim. **Proposição De Um Método De Planejamento E Gestão Estratégica De Clusters.** 2012. 190p. (Dissertação de Mestrado) – Universidade de São Paulo. Ribeirão Preto. Disponível em: < http://www.teses.usp.br/teses/disponiveis/96/96132/tde-06122012-154727/pt-br.php>. Acesso em: 20 fevereiro de 2016.

AMAZONAS, M. C. **Valor e meio ambiente: elementos para uma abordagem evolucionista. 2001, 267f.** (Tese de doutorado em Economia, área de concentração: Teoria Econômica) – Instituto de Economia, Universidade Estadual de Campinas. Campinas. Disponível em: < http://www.bibliotecadigital.unicamp.br/document/?code=vtls000220503>. Acesso em: 21 de fevereiro de 2016.

ANSAR Atif, FLYVBJERG Bent, BUDZIER Alexander, LUNN Daniel. **Should we build more large dams? The actual costs of hydropower megaproject development.** Disponível em: <http://www.sbs.ox.ac.uk/faculty-research/megaproject-management/publications-0/journal-articles-0/should-we-build-more-large-dams-actual-costs-hydropower-megaproject-development>. Acesso em: 21 de fevereiro de 2016.

AWAZU, Luís Alberto de Fischer. **A importância da sustentabilidade do pacto federativo no Brasil e sua relação com o desenvolvimento nacional.** 2012. 198p. (Dissertação de Mestrado) – Universidade de São Paulo. São Paulo. Disponível em: < http://www.teses.usp.br/teses/disponiveis/2/2134/tde-29082013-133748/pt-br.php>. Acesso em: 20 fevereiro de 2016.

Astah Documentação. Disponível em: <http://www.astah.net>. 2014. Acesso em: 01 novembro de 2014.

AYRES, R., SIMONIS, U. E. **Industrial Metabolism: Restructuring for Sustainable Development.** Disponível em: <http://archive.unu.edu/unupress/unupbooks/80841e/80841E00.htm >.Tokyo. The United Nations University. 1994. 390 p.

Bakkes, J.A., Van den Born, G.J., Helder, J.C., Swart, R. J., Hope, C.W., and Parker, J.D.E. **An Overview of Environmental Indicators: State of the Art and Perspectives.** UNEP/EATR.94-01; RIVM/402001001. Environmental Assessment Sub-Programme. UNEP. Nairobi. 1994.

BELLEN, Hans Michael Van. **Indicadores de sustentabilidade: uma análise comparativa 2.ed.** Rio de Janeiro: Fundação Getúlio Vargas. 2006.

BERGER, J. O.; INSUA, D. R. **Recent developments in Bayesian inference with applications in hydrology**. [S.l.]: Censiglio and Nazionale Delle Ricerche. 1996.

BERNARDO, J. M., SMITH, A. F. M.. **Bayesian Theory**. Nova Iorque. Wiley. 1994

BRASIL. **Política Nacional de Desenvolvimento Regional – PNDR.** Disponível em:

<http://www.integracao.gov.br/politica-nacional-de-desenvolvimento-regional-pndr>. Acessado em: 30 julho de 2014.

CALDERARO, I. F. N., DE ALMEIDA, F.M. **C-Language Functions for PostgreSQL in Modeling Software for Decision Making Through Indicators.** International Science Congress Association. v. 1. p. 1-15. IVC 2014. ISBN: 978-93-83520-30-5. Disponível em: <http://isca.net.co/vconference/subaccept.php>. Acesso em: 10 junho de 2014.

CALDERARO, I. F. N., DE ALMEIDA, F.M. **Software Development Based on Indicators.** Review Of Research. v. 3. p. 1-7. 2014. DOI:10.9780/2249-894X/352014/613. Disponível em: <http://ror.isrj.org/ArticleDetails.aspx?id=613>. Acesso em: 05 fevereiro de 2014.

CARLIN, Bradley P., LOUIS Thomas A. **Bayes and Empirical Bayes Methods for Data Analysis**, 2nd ed. Lodon. Chapman & Hall/CRC, A CRC Press Company. 2000. 434p.

CAVALCANTE, Fábio Robson Casara, GÓES, Silvia Bezerra de. **"Correlação entre desflorestamento e pecuária bovina no estado de Rondônia: Um estudo soba a perspectiva microregional"**. 2013 Disponível em: <http://www.ibeas.org.br/congresso/Trabalhos2013/VI-080.pdf>. Acesso em: 01 maio de 2016.

CHARNIAK, Eugene. **"Bayesians Networks without Tears"**. IA Magazine, 1991. 14f. Disponível em: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/918/836>. Acesso em: 01 novembro de 2012.

CASTILHO, E., GUTIERREZ J.. **"Expert Systems and Probabilistic Network Models".** Ed. Springer. 1997.

CASELLA, G., GEORGE, E. I.. **"Explaining the Gibbs sampler".** Am. Stat. 46p. 1992.

CodeBlocks Documentação. Disponível em: < http://www.codeblocks.org/features>. 2014. Acesso em: 01 novembro de 2014.

DARWICHE, Adan & Huang, Cecil. **"Inference in Belief Networks: A procedural guide"**. 1994. 39f. International Journal of Approximate Reasoning, Disponível em: < http://www.sciencedirect.com/science/article/pii/S0888613X96000692>. Acesso em: 01 novembro de 2012.

DINIZ, M. B.. JUNIOR, J. N. O.. NETO, N. T.. DINIZ, M. J. T.. **Causas do desmatamento da Amazônia: uma aplicação do teste de causalidade de Granger acerca das principais fontes de desmatamento nos municípios da Amazônia Legal brasileira**. 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-63512009000100006>. Acesso em: 01 de maio de 2016.

FRANZIN, Sergio F. L., DE ALMEIDA, Fabrício M., DE SOUZA Carlos H. M. **A Inovação e o Desenvolvimento Regional como Referência para Políticas públicas no Brasil.** 2014. 20f. Inter Science Place. Edição 29, volume 1, artigo nº 5, Junho. 2014. D.O.I:10.6020/1679-9844/2905. Disponível em: <http://www.interscienceplace.org/interscienceplace/article/view/300>. Acesso em: 15 agosto de 2014.

GERSTING, Judith L. **Fundamentos Matemáticos para a Ciência da Computação.** 4 ed. Rio de Janeiroeiro. LTC - Livros Técnicos e Científicos. Editora S. A. 2001. 533p.

GELMAN, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. **Bayesian Data Analysis**. London. Chapman & Hall/CRC, A CRC Press Company. 1995. 526p.

GEMAN, S.. GEMAN, D.. **"Stochastic relaxation, Gibbs distribution and Bayesian restoration of images"**. IEEE Transaction on Pattern Analysis and Machine Intelligence 6. 1984.

GUEDES, Gilleanes T. A. **UML 2 Uma Abordagem Prática**. São Paulo. Novatec Editora LTDA. 2009. 481p.

MPOG. **Guia Referencial Para Medição De Desempenho E Manual Para Construção De Indicadores.** Governo Federal. Ministério do Planejamento Orçamento e Gestão – MPOG. Brasil. Disponível em: < www.gespublica.gov.br/sites/default/files/documentos/guia_indicadores_jun2010.pdf>. Acessado em: 12 novembro de 2013.

HARDI, P., BARG, S. **Measuring Sustainable Development: Review of Current Practice.** Winnipeg. IISD. 1997.

HOSMER, D. W., LEMESHOW, S. **Applied logistic regression. In: Applied logistic regression.** [S.1.]. Wiley. 2000.

JENSEN, F. V. **Bayesian networks and decision graphs. statistics for engineering and information science**. Springer. v32. p34. 2001.

JR, Claudionei Nalle. **Desenvolvimento Regional E Políticas Públicas: O Caso Do Projeto Amanhã Da Companhia De Desenvolvimento Dos Vales Do São Francisco E Parnaíba.** 2006. 202p. (Dissertação de Mestrado) – Universidade de São Paulo. Ribeirão Preto. Disponível em: <http://www.teses.usp.br/teses/disponiveis/96/96132/tde-06022007-134845/pt-br.php >. Acesso em: 15 fevereiro de 2016.

JÚNIOR, Ademir Antonio Pereira. **Sistema Financeiro, Desenvolvimento Regional E Estado: A Regulamentação Jurídica Do Crédito Financeiro.** 2013. 193p. (Dissertação de Mestrado) – Universidade de São Paulo. São Paulo. Disponível em: < http://www.teses.usp.br/teses/disponiveis/2/2133/tde-27112013-153537/pt-br.php>. Acesso em: 18 fevereiro de 2016.

KAMPEL, Silvana Amaral, CÂMARA, Gilberto. **Análise Exploratória das Relações Espaciais do Desflorestamento Na Amazônia Legal Brasileira.** Disponível em: < http://www.dpi.inpe.br/gilberto/papers/silvana_gisbrasil2000.pdf>. Acesso em: 01 maio de 2016.

KRAMMA, Márcia Regina. **Análise dos indicadores de desenvolvimento sustentável no Brasil, usando a ferramenta painel de sustentabilidade.** 2009. 185p. (Dissertação de Mestrado) – Universidade Católica do Paraná. Curitiba. Disponível em: <http://indicadores.fecam.org.br/uploads/28/arquivos/4056_KRAMA_M_Indicadores_de_Sustentabilidade_no_Brasil_aplicando_o_Dashboard_of_Sustainablity.pdf>. Acesso em: 21 de fevereiro de 2016.

KUBO, Adriana Yumi, TENORIO, Marcelo Buscioli, FAVARIN Sidnei. **Rede Bayesiana Aplicada a Tomada de Decisão na Produção do Leite Bovino.** 2010. 14 f. (Artigo) - FATEC de Presidente Prudente. São Paulo. Disponível em: < http://www.revistasapere.inf.br/download/quarta/ADRIANA_MARCELO_SIDNEI.pdf >. Acesso em: 01 janeiro de 2013.

LUGER, George F. **Inteligência Artificial: Estruturas e Estratégias para a Solução de Problemas Complexos.** 4. ed. Porto Alegre. Bookman. 2004. 781p.

MOLDAN, B.; BILHARZ, S. (Eds.). **Sustainability indicators: report of the project on indicators of sustainable development.** Chichester: John Wiley & Sons Ltd. UK. 1997.

MULLER, Ione Novoa Jezler. **Infra-Estruturas De Apoio A Grandes Empreendimentos E As Alterações No Meio Ambiente.** 1994. 202p. (Dissertação de Mestrado) – Universidade de São Paulo. São Paulo. Disponível em:< http://www.teses.usp.br/teses/disponiveis/90/90131/tde-04112011-185846/pt-br.php >. Acesso em: 12 fevereiro de 2016.

NEAPOLITAN, R. E. et al. **Learning bayesian networks**. [S.l.]. Prentice Hall Upper Saddle River. 2004.

NEVES, Ana Carolina Marzullo. **Determinantes do Desmatamento na Mata Atlântica: Uma Análise Econômica**. 2006. 94p. (Dissertação de Mestrado) – Universidade Federal do Rio de Janeiro. Rio de Janeiro. Disponível em: < http://www.ie.ufrj.br/images/conjuntura/Gema_Dissertaes/AnaCarolinaMarzulloNeves_2006_IE_Determinantes_do_desmatamento_na_mata_atlantica.pdf>. Acesso em: 22 abril de 2016.

NORTH, D. **Teoria da Localização e Crescimento Econômico Regional.** In: SCHWARTZMANN, J. (Org.). Economia Regional e Urbana: Textos Escolhidos. Belo Horizonte: UFMG, 1977.

OCDE. **Manual de Oslo: diretrizes para coleta e interpretação de dados sobre inovoação.** 3. ed. , São Paulo. Finep. 2005. 136p.

OLIVEIRA, Aparecida Antonia. **Políticas Ambientais e Desenvolvimento Regional: A Perspectiva do Pensamento Institucionalista Evolucionário.** 2012. 282f. (Tese de Doutorado em Economia) - Universidade Federal do Rio Grande do Sul. Porto Alegre.
Disponível em: <http://www.lume.ufrgs.br/handle/10183/69998>. Acesso em: 02 fevereiro de 2013.

OMG Unified Modeling Language TM (OMG UML) version 2.5. Disponível em: <http://www.omg.org/spec/UML/2.5/>. 2015. Acessado em: 04 de abril de 2016

OREGON STATE UNIVERSITY, USA. **Integrative Dam Assessment Modeling - IDAM. Modelling the Effects of Dams.** 2012. Disponível em: <http://rivers.bee.oregonstate.edu/integrative-dam-assessment-modelling-idam>. Acessado em: 05 julho de 2012.

PEARCE, D. **An Intellectual History of Environmental Economics. Annual Reviews EnergyEnvironment**., [s.n.], n. 27, p. 57-81, 2002.

PEARL, J.. **Fusion, propagation, and structuring in belief networks**. Journal of Artificial Intelligence, v.29 , p.241-288, 1986

PEREIRA, Claudio Robinson Tapié. **Sistema de Tomada de Decisão Para Compra e Venda de Ativos Financeiros Utilizando Lógica Fuzzy**. 2008. 128f. (Dissertação de Mestrado em Engenharia Elétrica) - Escola Politécnica da Universidade de São Paulo. São Paulo. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3142/tde-06112008-104532/publico/SISTOMDEC_CLAUDIO_2008_FINAL_Ed_Revisada.pdf>. Acesso em: 01 janeiro de 2013.

PEREIRA, Rodrigo Mendes. GÓES, Geraldo Sandoval. **O Desmatamento Amazônico e o Ciclo Econômico no Brasil**. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/5563/1/BRU_n07_desmatamento.pdf>. Acesso em: 22 abril de 2016.

PgAdmin Documentação. Disponível em: <http://www.pgadmin.org>. 2014. Acesso em: 01 novembro de 2014.

PgModeler Documentação. Disponível em: <http://www.pgmodeler.com.br/api/0.8.0-alpha1/>. 2014. Acesso em: 01 novembro de 2014.

PONTES. Antonio Carlos Fonseca. **Análise De Variância Multivariada Com A Utilização De Testes Não-Paramétricos E Componentes Principais Baseados Em Matrizes De Postos.** 2005. 117p. (Tese de Doutorado) – Universidade de São Paulo. Piracicaba. Disponível em:<www.lce.esalq.usp.br/tadeu/AntonioPontes_tese.pdf>. Acesso em: 12 fevereiro de 2016.

PostgreSQL 9.3.5 Documentação. Disponível em: < http://www.postgresql.org/docs/current/static/release.html>. 2014. Acesso em: 01 novembro de 2014.

PRESCOTT-ALLEN, R. **Assessing Progress toward Sustainability: The System Assessment Method illustrated by the Wellbeing of Nations**. Cambridge: IUCN, 1999.

PRESCOTT-ALLEN, R. **Barometer of Sustainability: Measuring and communicating wellbeing and sustainable development**. Cambridge: IUCN, 1997.

PRESCOTT-ALLEN, R. (1997). **Barometer of sustainability. In: Moldan, Bedrich; Bilharz, Suzanne. (1997) Sustainability indicators: report of the project on indicators of sustainable development**. Chicester: John Wiley & Sons Ltd.

PRESSMAN, Roger S.; LOWE, David. **Engenharia WEB.** Rio de Janeiroeiro. LTC - Livros Técnicos e Científicos Editora S. A. 2009. 408 p.

PRODES. **Monitoramento da Floresta Amazônica Brasileira por Satélite**. Disponível em: <http://www.obt.inpe.br/prodes/index.php>. Acessado em: 30 agosto de 2015.

ROSEMANN, Douglas. **Sistema Tutor Inteligente Utilizando Redes Bayesianas – Estudo De Caso De Uma Disciplina Do Curso De Ciência Da Computação.** 2015. 205p. (Dissertação de Mestrado) – Universidade do Vale do Itajaí. Itajaí. Disponível em:<siaibib01.univali.br/pdf/Douglas%20Rosemann.pdf >. Acesso em: 02 fevereiro de 2016.

RUSSEL, S. J.. NORVIG, P.. **Inteligência Artificial**. Tradução da 2a Edição. Ed.Campus, 2004

SANTOS, Renato Prado dos. **Os Principais fatores do desmatamento na Amazônia (2002-2007) – uma análise econométrica e espacial.** 2010. 129p. (Dissertação de Mestrado) – Universidade de Brasília. Brasília. Disponível em:< http://repositorio.unb.br/bitstream/10482/6592/1/2010_RenatoPradodosSantos.pdf>. Acesso em: 22 abril de 2016.

SUZIĆ, R.. **Generic Representation of Military Organisation and Military Behaviour: UML and Bayesian Networks. In Proceedings of the NATO RTO Symposium on C3I and M&S Interoperability.** 2003. 10p. Analya. Turkey. Disponível em: <http://www.csc.kth.se/~rsu/>. Acesso em: 07 de abril de 2016.

SUZIĆ, R.. **Stochastic Multi-Agent Plan Recognition, Knowledge Representation and Simulations for Efficient Decision Making**. 2006. 121p. (Tese de Doutorado). Royal Institute of Technology (KTH). Stockholm. Sweden. Disponível em: <http://www.csc.kth.se/~rsu/> . Acesso em: 07 de abril de 2016.

Ubuntu GNOME 14.04.02 LTS. Disponível em: < https://wiki.ubuntu.com/UbuntuGNOME/GetUbuntuGNOME >.2014. Acessado em: 02 julho de 2014.

UML 2.0 Documentação. Disponível em:< http://www.uml.org/#UML2.0>. Acesso em: 07 maio de 2015.

WETTIG, H., PERNESTAL, A., SILANDER, T., NYBERG, MATTIAS**. A Bayesian Approach to Learning in Fault Isolation.** Diposnível em: http://www.cs.uu.nl/groups/DSS/UAI08-workshop/. Acesso em: 07 de abril de 2016.